# Approximation and Bounds in Discrete Event Dynamic Programming

ALAIN HAURIE, MEMBER, IEEE, AND PIERRE L'ECUYER, MEMBER, IEEE

*Abstract*—**This paper presents a general dynamic programming algorithm for the solution of optimal stochastic control problems concerning a class of discrete event systems. The emphasis is put on the numerical technique used for the approximation of the solution of the dynamic programming equation.**

**This approach can be efficiently used for the solution of optimal control problems concerning Markov renewal processes. This is illustrated on a group preventive replacement model generalizing an earlier work of the authors.**

## I. Introduction

THIS paper deals with the computation of optimal control laws for a class of discrete-event systems. These systems are typical of the modeling of queueing or maintenance problems, and we shall illustrate the computational techniques presented in this paper, with a preventive replacement model.

In a typical preventive replacement problem (see [9], [10]), the state of a component is described by its age or by an indication of failure. Replacement of the component restores its age at a zero value. In a discrete event formulation of the problem, the system is observed at discrete random times. At any of these times, also called intervention times, the controller observes the state of the system and chooses an action in an admissible set. Associated with the state and action, a probability law is governing the generation of the next intervention time and observed state.

As the age is a continuous variable, the system can tentatively be modeled as a Markov decision process, provided that the state and action sets be defined as Borelian spaces. Bertsekas and Shreve [5] have proposed a rather complete theory of infinite horizon dynamic programming with such general state and action spaces. However, their theory is based on a discrete-time formulation with a constant discount factor.

In a discrete-event system, the times of intervention are random, so the discount factor will not remain constant from one stage to the other. However, for a large class of problems this discount factor can be expressed as a function of the current state of the system (which may include the current time).

Various authors [11], [21], [25] have already studied infinite horizon multistage decision processes with state-dependent discount factors, usually assuming that the integral of the stochastic transition kernel was uniformly bounded away from 1. However, for many discrete event systems, this assumption does not hold.

Whittle [29] obtained a condition (the bridging condition) for a Markov decision process with nonnegative costs to enjoy regularity properties implying the limit of $n$-stage optimal policies to be optimal.

Finally, a generalization of the Bertsekas–Shreve dynamic programming formalism has been achieved in [14] and [15], without the latter uniform boundedness assumption. The present paper is concerned with the numerical solution of the general dynamic programming equation obtained in [15].

The simplest approach for the implementation of a numerical technique would be to discretize the time, state, and action domains, and then get back to a standard problem of a discrete (although very large) Markovian decision process. This approach is illustrated in [10], in the setting of an optimal group replacement problem.

A more satisfactory approach consists of using numerical approximation techniques for the solution of the dynamic programming equation of the original problem. This is the approach followed in this paper.

The paper is organized as follows. In Section II we review briefly the most relevant results of [15] which concern a general multistage decision process; in Section III we present the use of approximation techniques for the solution of the dynamic programming equation. In Section IV we see that our model encompasses the Markov-renewal decision processes with Borelian state and action spaces. Finally in Section V we use the approach for the numerical solution of a complex maintenance problem.

## II. A Class of Multistage Decision Processes with State-Dependent Discount Factor

In this section we present some results which extend the dynamic programming theory of [5] to the case of a discounted Markov decision process with a state-dependent discount factor. The proof of these results is given in [14] and [15].

We consider a system with state $x$ in $X$, a given Borelian space. At each decision time, a controller chooses an action $a$ in $A$, also a Borelian space. A state-dependent constraint on the action set is defined by an analytic subset $\Gamma$ of $X \times A$, called the set of admissible state-action couples, such that

$$A(x) \triangleq \{a \in A : (x, a) \in \Gamma\} \neq \phi \qquad \forall x \in X. \quad (1)$$

The dynamics of the system is described by a stochastic kernel $Q$, which is a family $\{Q(\cdot \mid x, a), (x, a) \in X \times A\}$ of probability laws on $X$. Viewed as a functional $Q: X \times A \to \mathcal{P}(X)$, $Q$ is assumed to be Borel-measurable, where the set $\mathcal{P}(X)$ of probability measures on $X$ is endowed with the topology of weak convergence. The one-stage cost function is a lower semi-analytic function $g: \Gamma \to \boldsymbol{R}$. At stage $n$, if the system is in state $x_n$ and the controller picks action $a_n$ in $A(x_n)$, then a cost $g(x_n, a_n)$ is incurred for that stage, and the state at stage $n+1$ is generated randomly according to the probability measure $Q(\cdot \mid x_n, a_n)$. The cost incurred at stage $n$ is discounted to a given origin by a discount factor $\beta(x_n)$, where $\beta: X \to (0, 1]$ is a Borel measurable discounting function. The system is assumed to operate over an infinite time horizon (i.e., an infinite number of stages).

Manuscript received July 13, 1984; revised August 12, 1985. Paper recommended by Past Associate Editor, J. Walrand. This paper is based on a prior submission of October 10, 1983. This work was supported by SSHRC-Canada under Grant 410-83-1012, NSERC-Canada under Grant A4952, FCAC-Québec under Grant EQ-0428, and NSERC-Canada under Grant A5463.

A. Haurie is with GERAD, Ecole des Hautes Etudes Commerciales, Montréal, P.Q., Canada.

P. L'Ecuyer is with the Département d'Informatique, Université Laval, Cité Universitaire, Québec, Canada.

IEEE Log Number 8406585.

*Remark 2.1:* Notice that this definition of the discounting factor implies that the current time $t$ is included as a component of the state $x$. This is particularly true for the classical Markov renewal decision process with discounting [7] which is encompassed by the present model. The inclusion of time in the state variable permits also the consideration in a unique framework of stationary (time homogeneous) as well as nonstationary systems.

The discounted cost for stage $n$ is denoted by

$$c_n \triangleq \beta(x_n) g(x_n, a_n). \tag{2}$$

We assume that the controller picks an action at each stage by using an admissible stationary feedback control law defined as a function $\mu: X \to A$ which is universally measurable and satisfies

$$\mu(x) \in A(x) \ \forall\, x \in X. \tag{3}$$

Associated with any initial state $x$ and control law, there is a uniquely defined probability measure $P_{\mu,x}$ (see [14], [15]) on the set $H$ of infinite sequences $(x_0, a_0, x_1, a_1, \cdots)$ where

$$(x_i, a_i) \in X \times A \qquad \text{for } i = 0, 1, \cdots \text{ and } x_0 = x.$$

*Definition 2.1:* A multistage decision process with state-dependent discount factor is well defined by $(X, A, \Gamma, Q, g, \beta)$ if, for any initial state $x$ in $X$, the series

$$\sum_{i=0}^{\infty} c_i$$

is well defined (finite or infinite) $P_{\mu,x}$ almost everywhere and the integral

$$\int_H \left( \sum_{i=0}^{\infty} c_i \right) dP_{\mu,x} \tag{4}$$

is also well defined.

We introduce, when they exist, the values

$$V_\mu(x) \triangleq \frac{1}{\beta(x)} \int_H \left( \sum_{i=0}^{\infty} c_i \right) dP_{\mu,x}$$

and

$$V^*(x) \triangleq \inf_{\mu \in \mathfrak{U}} V_\mu(x)$$

where $\mathfrak{U}$ is the set of admissible control laws.

A control law $\mu$ in $\mathfrak{U}$ is said to be optimal (respectively, $\epsilon$-optimal) if $V_\mu(x) = V^*(x)$ (respectively, $V_\mu(x) \le V^*(x) + \epsilon$) for all $x$ in $X$.

*Remark 2.2:* An *a priori* more general definition of a control law could be used, with memory and a random determination of the action at a given state $x_n$. However, it could be shown [14], [15] that one can without loss of generality restrict the analysis to this simpler class of control laws.

The single important difference between this class of systems and the one thoroughly analyzed in [5] stems from the consideration of a discount factor defined as a function $\beta(x_n)$ instead of $\beta^n$, with $\beta \in (0, 1)$.

In the latter case, the condition that $\beta$ be in $(0, 1)$ induces a geometrically decreasing sequence of discount factors.

When the discount factor depends on $x$, one has to consider the expected discount factor from stage $n+1$ to stage $n$ which is given by

$$\alpha(x_n, a_n) = \frac{1}{\beta(x_n)} \int_X \beta(x_{n+1}) Q(dx_{n+1} | x_n, a_n) \tag{5}$$

and we will introduce two versions of the model depending on the particular assumptions made on the functions $g$, $\alpha$, and $Q$.

*Assumption 1.1:* There exist $\alpha_1$, $g_0 \le 0$, $g_1 \ge 0$ such that

$$\forall\, (x, a) \in \Gamma, \begin{cases} \alpha(x, a) \le \alpha_1 < 1 & (6) \\ g_0 \le g(x, a) \le g_1. & (7) \end{cases}$$

A system which satisfies (6), (7) will be called a "$C$-system" where $C$ means "contracting."

Another class of systems is associated with the next assumption, and many of them are not $C$-systems.

*Assumption 1.2:* There exists a feedback control law $\tilde{\mu}$ and real numbers $\delta_1$, $K_1$, $K_2$, $g_1$ such that $g_1 \ge 0$ and

$$\forall\, x \in X \qquad \alpha(x, \tilde{\mu}(x)) \le \delta_1 < 1 \tag{8}$$

$$K_1 + K_2 > 0 \tag{9}$$

$$\forall\, (x, a) \in \Gamma \qquad K_1 + K_2 \alpha(x, a) \le g(x, a) \tag{10}$$

$$\forall\, x \in X \qquad g(x, \tilde{\mu}(x)) \le g_1 \tag{11}$$

$$\forall\, (x, a) \in \Gamma \qquad \int_X 1_{\beta(x') > \beta(x)} Q(dx' | x, a) = 0. \tag{12}$$

A system which satisfies (8)–(12) will be called an "$LC$-system" where $LC$ stands for "locally contracting."

The dynamic programming approach summarized in the two forthcoming theorems will be valid under both assumptions.

Let $\mathfrak{B}_0$ be the Banach space of bounded functionals $V: X \to (-\infty, \infty)$ endowed with the norm $\|V\| = \sup_{x \in X} |V(x)|$, $\mathfrak{B}_1$ the subspace of lower semianalytic functionals $V$ in $\mathfrak{B}_0$ which are also universally measurable, and finally $\mathfrak{B}_2$ the closed subset of $\mathfrak{B}_1$ defined as follows:

$$\mathfrak{B}_2 = \begin{cases} \left\{ V \in \mathfrak{B}_1 : \frac{g_0}{1-\alpha_1} \le V \le \frac{g_1}{1-\alpha_1} \right\} \\ \qquad \text{for version } C \\ \left\{ V \in \mathfrak{B}_1 : K_1 + \min\,(0, K_2) \le V \le \frac{g_1}{1-\delta_1} \right\} \\ \qquad \text{for version } LC. \end{cases}$$

For any $V$ in $\mathfrak{B}_1$, we introduce

$$H(V)(x, a) \triangleq g(x, a) + \frac{1}{\beta(x)} \int_X \beta(x') V(x') Q(dx' | x, a) \qquad \forall\, (x, a) \in \Gamma \tag{13}$$

$$T_\mu(V)(x) \triangleq H(V)(x, \mu(x)) \qquad \forall\, x \in X, \ \forall\, \mu \in \mathfrak{U} \tag{14}$$

$$T(V)(x) \triangleq \inf_{a \in A(x)} H(V)(x, a) \qquad \forall\, x \in X. \tag{15}$$

The following two theorems summarize the dynamic programming approach.

*Theorem 2.1:* Let $V$ be a function belonging to $\mathfrak{B}_2$. Then the following holds for versions $C$ and $LC$ of the model.

a) $T(V) = V$ iff $V = V^*$.
b) If $T(V) \le V$, then $V^* \le V$.
c) If $T(V) \ge V$, then $V^* \ge V$.
d) $\lim_{n \to \infty} \| T^n(V) - V^* \| = 0$.
e) $V^* \in \mathfrak{B}_2$.
f) A control law $\mu$ is optimal iff $T_\mu(V^*) = V^*$.
g) A control law $\mu$ is optimal iff $T(V_\mu) = V_\mu \in \mathfrak{B}_2$.

*Remark 2.3:* This theorem gives a set of optimality conditions a), f), g), as well as the basis for a dynamic programming algorithm of successive approximations d) and properties permitting one to bound the optimal value function b), c).

This theorem proved in [14],[15] is complemented by the following, also proved in the same references.

*Theorem 2.2:* The following holds for both versions, $C$ and $LC$, of the model.

a) There exists an optimal control law $\mu$ iff the $\inf_{a \in A(x)} H(V^*)(x, a)$ is attained for all $x$ in $X$.

b) Let $V$ be any function in $\mathfrak{B}_2$. If for some integer $n_0$ the sets

$$U_n(V)(x, \lambda) \triangleq \{ a \in A(x) | H(T^n(V))(x, a) \le \lambda \}$$

are compact for any integer $n$ greater than $n_0$, then there exists a sequence $\{\mu_n\}_{n \in N}$ of control laws such that

$$T_{\mu_n}(T^n(V)) = T^{n+1}(V) \qquad \forall\, n \geq n_0,$$

and there exists a control law $\mu$ which is a pointwise limit of $\{\mu_n\}_{n\in N}$. This control law $\mu$ is also an optimal control law.

*Remark 2.4:* The conditions for Theorem 2.2 b) are satisfied, in particular, if each set $A(x)$ is finite or if each $A(x)$ is compact, the product $g \cdot \beta$ and $V$ are lower semicontinuous and $Q$ is continuous on $X \times A$.

Notice also that b) is a constructive proof of the existence of an optimal control law.

## III. Approximation Techniques for the Solution of the DP Equation

One of the last "frontiers" in the theory and applications of dynamic programming lies in the numerical solution of the DP equation by using approximation techniques. For example, Rishel [24] has proposed a value iteration method for the computation of the optimal control for a jump process describing a group preventive replacement problem. However, to be implemented, this approach would necessitate the use of approximation techniques in the computation of the value function at each iteration.

In this section we introduce a general algorithm based on a value iteration technique with approximate computation of the value function at each step. This algorithm is adapted to the DP equation given in Section II, and generalizes most of the approaches proposed in the OR literature for the computation of approximations and bounds in dynamic programs.

The value iteration technique consists of applying the DP operator $T$ repeatedly, until some convergence test has been passed. McQueen [16], Denardo [7], and Porteus [19], [20] have obtained bounds for the norms $\|T^n(V) - V^*\|$ and $\|V_\mu - V^*\|$ when $T$ is applied exactly at each iteration. These bounds converge geometrically to 0 as $n$ tends to infinity.

When the state space is infinite, it is necessary to use an approximate computation of $T(V)$ on $X$. A natural approach consists of partitioning the set $X$ into a finite class of subsets, selecting a representative state in each subset, and defining an approximate finite state model. Bellman and Dreyfus [2] have proposed the approach, Fox [8] has given conditions for the convergence of the value function to $V^*$. A similar scheme was proposed by Bertsekas [4] with a discretization of both the state and the action spaces. Whitt [28] extended the approach to the general comparison between dynamic programs. Hinderer [12], Langen [13], and several others have also contributed to the mathematical theory of the convergence of a sequence of dynamic programs. Typically, these authors obtain bounds for the norms $\|V^* - V_m\|$ and $\|V^* - V_\mu\|$. Here $V_m$ is the optimal value function for the approximate model which is extended to the whole state space $X$ as a function that is constant on each subset of the partition, and $V_\mu$ is the value-function obtained on $X$ when one uses the optimal policy $\mu$ of the approximate model, extended to $X$ by a function which is constant on each subset of the partition.

The approximation of the optimal value function $V^*$ on $X$ by a piecewise constant function could be advantageously replaced by more sophisticated schemes like polynomial approximation, spline interpolation or approximation, finite element methods, etc. Daniel [6] and Morin [18] have advocated such an approach.

Based on the dynamic programming equations summarized by Theorems 2.1 and 2.2, a general algorithm can be designed for the approximate computation of the optimal cost-to-go function. This algorithm uses bounds on $V^*$ defined by the next theorem whose proof is given in the Appendix.

For any function $V$ in $\mathfrak{B}_1$, we denote

$$V^-(x) \triangleq \max\,(0,\ V(x)).$$

*Theorem 3.1:* Let $\alpha_1$ satisfy (6) and $n_0 = 1$ for version $C$. Let $\alpha_1 \in (0, 1)$ and $n_0$ be the smallest integer larger than

$$\frac{\dfrac{g_1}{1-\delta_1} - K_1 - \min\,(0,\ K_2)}{\alpha_1(K_1 + K_2)}$$

for version $LC$ of the model.

Consider two functions $V$ and $V_1$ in $\mathfrak{B}_2$, and two numbers $\delta^-$, $\delta^+$ in $\boldsymbol{R}_+$ such that

$$-\delta^- \leq T(V) - V_1 \leq \delta^+.$$

Define

$$\epsilon^+ \triangleq n_0\delta^+ + (n_0 - 1)\|(V_1 - V)^+\| \tag{16}$$

$$\epsilon^- \triangleq n_0\delta^- + (n_0 - 1)\|(V - V_1)^+\|. \tag{17}$$

Then the following holds.

a) $V^*$

$$\geq V_1 - \epsilon^- - \frac{\alpha_1}{1-\alpha_1}\,\|(V - V_1 + \epsilon^-)^+\| \tag{18}$$

$$\leq V_1 + \epsilon^+ + \frac{\alpha_1}{1-\alpha_1}\,\|(V_1 + \epsilon^+ - V)^+\|. \tag{19}$$

b) For any $\epsilon_0 > \delta^+$ there exists a control law $\mu$ such that

$$T_\mu(V) \leq T(V) + \epsilon_0 - \delta^+ \leq V_1 + \epsilon_0. \tag{20}$$

If there exists $\alpha \in [0, 1)$ such that

$$T_\mu(V_3) - T_\mu(V_2) \leq \alpha\|V_3 - V_2\| \tag{21}$$

for any pair $(V_2, V_3) \in \mathfrak{B}_1^2$ satisfying $V_3 \geq V_2$, then

$$V^* \leq V_\mu \leq V^* + \epsilon^- + \frac{\alpha_1}{1-\alpha_1}\,\|(V - V_1 + \epsilon^-)^+\| + \epsilon_0 + \frac{\alpha}{1-\alpha}\,\|(V_1 + \epsilon_0 - V)^+\|. \quad \blacksquare \tag{22}$$

This theorem is complemented by the following one, due originally to MacQueen [17], which permits the elimination of nonoptimal actions. Various other elimination procedures are discussed in [5], [19], [20], [22], [27], and [28].

*Theorem 3.2:* If $\underline{V}$ and $\bar{V}$ are two functions in $\mathfrak{B}_2$ such that

$$\underline{V} \leq V^* \leq \bar{V}$$

and, if $a \in A(x)$ is such that

$$H(\underline{V})(x,\ a) > \bar{V}(x),$$

then $a \notin A^*(x)$.

*Proof:*

$$H(\underline{V})(x,\ a) > \bar{V}(x) \Rightarrow H(V^*)(x,\ a) > V^*(x) \Rightarrow a \notin A^*(x). \quad \blacksquare$$

These two theorems suggest the following algorithm.

### Algorithm

① Set $\alpha_1$ and $n_0$ as in Theorem 3.1.
Set $\tilde{\mathfrak{B}} := \mathfrak{B}_2$.
Choose an upper bound for the total number of iterations.
Choose $\epsilon > 0$.
Take any function $V$ in $\tilde{\mathfrak{B}}$ as an initial guess of $V^*$.

② Compute $T(V)$ at a finite number of points in $X$.
Define $V_1$ in $\tilde{\mathfrak{B}}$ as an approximation of $T(V)$ on $X$.

③ Compute $\delta^-$, $\delta^+$ in $\boldsymbol{R}_+$ such that
$-\delta^- \leq T(V) - V_1 \leq \delta^+$
and compute $\epsilon^-$ and $\epsilon^-$ as in (16), (17).
The inequalities (18), (19) determine bounds for $V^*$.
*Stopping Rule 1:* Stop if the difference between the two bounds (lower and upper) is smaller than $\epsilon$.

④ Find a control law $\mu$ and $\epsilon_0 \geq 0$ such that

$$T_\mu(V) \leq V_1 + \epsilon_0.$$

If there exists $\alpha \in [0, 1)$ such that

$$T_\mu(V_3) - T_\mu(V_2) \le \alpha \| V_3 - V_2 \|$$

for any pair $(V_2, V_3) \in \mathcal{B}_1^2$ and if

$$\epsilon^- + \frac{\alpha_1}{1-\alpha_1} \| (V - V_1 + \epsilon^-)^+ \| + \epsilon_0 + \frac{\alpha}{1-\alpha} \| (V_1 + \epsilon_0 - V)^+ \| \le \epsilon \quad (23)$$

then $\mu$ is a control law which is $\epsilon$-optimal.

*Stopping Rule 2:* Stop when one has obtained an $\epsilon$-optimal control law.

⑤ *Stopping Rule 3:* Stop when the maximal number of iterations has been attained.

⑥ Set

$$\underline{V}(x) := \max \left( \underline{V}(x),\ V_1(x) - \epsilon^- - \frac{\alpha_1}{1-\alpha_1} \| (V - V_1 + \epsilon^-)^+ \| \right)$$

$$\bar{V}(x) := \min \left( \bar{V}(x),\ V_1(x) + \epsilon^+ + \frac{\alpha_1}{1-\alpha_1} \| (V_1 + \epsilon^+ - V)^+ \| \right)$$

$$\bar{\mathcal{B}} := \{ V \in \bar{\mathcal{B}} : \underline{V} \le V \le \bar{V} \}$$

$$V := V_1$$

⑦ Go to ②.

The following theorem whose proof is also given in the Appendix establishes the convergence of the algorithm.

*Theorem 3.3:* a) Let $V_0$ in $\mathcal{B}_2$ and a sequence

$$\{(\delta_n^-,\ \delta_n^+,\ V_n)\}_{n \in N} \text{ in } \boldsymbol{R}_+ \times \boldsymbol{R}_+ \times \mathcal{B}_2$$

be such that

$$\lim_{n \to \infty} \delta_n^- = \lim_{n \to \infty} \delta_n^- = 0$$

with

$$-\delta_n^- \le T(V_{n-1}) - V_n \le \delta_n^+ \qquad \forall\, n \in N.$$

Then

$$\lim_{n \to \infty} \| V_n - V^* \| = 0.$$

b) For the version $C$ of the model, if the sequences of $\delta^-$, $\delta^+$, and $\epsilon_0$ values obtained in steps ③ and ④ of the algorithm converge to 0, then for any $\epsilon > 0$ an $\epsilon$-optimal control law is obtained in a finite number of iterations. ■

*Remarks 3.1:* a) After step ⑥ of the algorithm, an elimination of nonoptimal actions can be done by using Theorem 3.2. Also the parameters $\alpha_1$ and $n_0$ could be reevaluated.

b) Step ② can be repeated any number of times before going to step ③, resetting $V := V_1$ after each repetition.

c) One of the originalities of this algorithm is that it provides bounds which take into account simultaneously the errors due to the approximation, at each step of the DP procedure, and the errors due to the fact that only a finite number of iterations are made.

d) The algorithm is very flexible. One can choose any method of approximation and even change the method from iteration to iteration.

e) In the implementation of the algorithm, a nontrivial task is to obtain values for $\delta^-$, $\delta^+$, and $\epsilon_0$. Obviously, finite values always exist. At worst, one can take $\bar{V} - \underline{V}$, which is, of course, quite pessimistic. For most practical situations, better bounds for the approximation error can be obtained. For instance, if $T(V)$ and $V_1$ are monotonous functions, and if $T(V)(x)$ can be computed at any given $x$ with negligible error, one can compute $T(V)$ on a finite grid and interpolate to define $V_1$. For example, if $X$ is the real interval $[a, b]$, one computes $T(V)$ at the $n$ points $a = x_1 < x_2 < \cdots < x_n = b$, defines

$$V_1(x) = \begin{cases} T(V)(x) & \text{if } x = x_i,\ i = 1, \cdots, n \\ \dfrac{T(V)(x_{i+1}) + T(V)(x_i)}{2} & \text{if } x_i < x < x_{i+1} \end{cases}$$

and set

$$\delta^- = \delta^+ = \max_{1 \le i \le n-1} \left| \frac{T(V)(x_{i+1}) - T(V)(x_i)}{2} \right|.$$

This also generalizes naturally to multidimensional state spaces.

Unfortunately, the main drawback of these "guaranteed" bounds on $T(V) - V_1$ is that they provide bounds on $V_*$ that are so conservative as to be useless in most practical situations. This same drawback also applies to the bounds proposed in [20] (see [3]).

Instead of computing guaranteed bounds, an alternative approach could be to estimate the real approximation error, and obtain estimate bounds on $T(V) - V_1$.

For instance, one way to estimate $\delta^-$ and $\delta^+$, when $T(V)$ is reasonably smooth, is to recompute $T(V)$ at a very large number of new points and compute the real approximation error at these points. If these points are well chosen, numerous, and if $T(V) - V$ behaves reasonably, then the smallest and largest of these errors can be taken as estimates of $\delta^-$ and $\delta^+$, respectively.

## IV. A Class of Markov-Renewal Decision Processes

We consider in this section an important subclass of systems which can be modeled as discounted Markov decision processes with a state-dependent discount factor. These systems, called Markov-renewal decision processes (MRDP) have a state set $X$ defined as a Cartesian product

$$X \triangleq \boldsymbol{R}^+ \times S$$

where $S$ is a given Borel space.

At stage $n$ the state of the system will be thus represented by a pair $x_n = (t_n, s_n)$, where $t_n$ corresponds to the time of occurrence of stage $n$ and $s_n$ will represent the "physical" state of the system.

Furthermore, for this class of systems, the discount factor $\beta(x_n)$ is defined as

$$\beta(x_n) \triangleq e^{-\rho t_n}$$

where $\rho$ is a positive (continuous) discount rate.

Let $A$ be the action space and $A(x_n)$ the set of admissible actions when the system is in state $x_n = (t_n, s_n)$, which are defined as in Section II.

The system dynamics can be described as follows: at stage 0, the initial state $x_0 = (t_0, s_0)$ is given; at any stage $n$, the controller observes the state $x_n = (t_n, s_n)$ and chooses an action $a_n$ in $A(x_n)$. Then the time $t_{n+1}$ of the next stage, with the next physical state $s_{n+1}$ are determined as

$$t_{n+1} = t_n + \zeta$$

$$s_{n+1} = s$$

where the pair $(\zeta, s)$ is generated randomly according to the probability measure $\tilde{Q}(\cdot \mid t_n, s_n, a_n)$. Here, the stochastic kernel $\tilde{Q}$ is a family $\{\tilde{Q}(\cdot \mid x, a), (x, a) \in X \times A\}$ of probability laws on $[0, \infty) \times S$.

Such a system is called a semi-Markov decision process if the measure $\tilde{Q}$ is such that almost surely $\zeta$ is nonzero. The cost of transition from stage $n$ to stage $n + 1$ is given by

$$g(x_n,\ a_n) = g(t_n,\ s_n,\ a_n).$$

The important subclass of homogeneous Markov-renewal decision processes is obtained when $A$, $\tilde{Q}$, and $g$ are not dependent on the time component $t_n$ of the state $x_n = (t_n, s_n)$. Then the control law $\mu$ can be a mapping from $S$ into $A$ such that

$$\forall\ s \in S \qquad \mu(s) \in A(s).$$

The Markov renewal decision process will define a $C$-system or a $LC$-system if it satisfies Assumption 1 or Assumption 2. Notice that, for this model we have

$$\alpha(t,\ s,\ a) = \int_{R^+} e^{-\rho\zeta} \tilde{Q}(d\zeta,\ S|s,\ a).$$

and

$$H(V)(t,\ s,\ a) = g(s,\ a) + \int_{[0,\ \infty)\times S} e^{-\rho\zeta} V(t+\zeta,\ s')\tilde{Q}(d\zeta,\ ds'|s,\ a).$$

*Remark 4.1:* If $V(t, s)$ is independent of $t$, then $H(V)(t, s, a)$ and $T(V)(t, s)$ are also independent of $t$. This shows that, for the homogeneous model, the time could be eliminated from the state description, and the class of functions $V$, $H(V)$, $T(V)$ would now have only $s$ and $a$ for arguments.

## V. Example: A Multicomponent System with Identical Elements

Consider a system comprised of $m$ identical and stochastically independent components, each having a known and nondecreasing failure rate $\lambda(t)$, a lifetime distribution function $F(t) = \exp(-\int_0^t \lambda(s)ds)$, and a survival function $\bar{F}(t) = 1 - F(t)$. Whenever a component fails, the repairman is instantly informed, must replace it at once by a new one (emergency replacement), and may replace at the same time any number of working components (preventive replacement). He can also halt the system at any moment and replace preventively any number of working components. All the replacement durations are assumed to be negligible.

The cost of an intervention is composed of a fixed cost $c_i$, and a replacement cost $c_r$ for each component replaced. A failure cost $c_f$ is also incurred each time a component fails, and all the costs are discounted at rate $\rho < 0$.

This generalizes the model studied in [1], [10], and [24], since neither the failure cost $c_f$ nor the possibility to intervene at any moment were considered then.

The system is observed whenever a component fails or a preventive replacement is performed. These observation times are also the decision points, at each of which an action is taken. An action is a couple $(l, d)$ where $l \in \{1, \cdots, m\}$ is a number of components to replace, and $d \in [0, \infty]$ is a time interval until the next planned preventive replacement. The state of a component is given by its age, where by convention the age of a failed component is $\infty$.

Owing to the nondecreasing failure rate, and since the components are identical, the $l$ components to be replaced are certainly the oldest ones. Hence, at any decision point, the oldest component is always to be replaced, and it suffices to consider only the states (ages) of the $m - 1$ others, in decreasing order.

The state and action spaces are defined as $X = [0, \infty) \times S$ where

$$S = \{(x_1,\ x_2,\ \cdots,\ x_{m-1}) \in R^{m-1} | x_1 \geq x_2 \geq \cdots \geq x_{m-1} \geq 0\}$$

and

$$A = \{1,\ \cdots,\ m\} \times [0,\ \infty]$$

respectively. One can easily define[1] $Q$ (using $\lambda$) and $g$, set $\Gamma = X \times A$, and verify that $(X, A, \Gamma, Q, g, \rho)$ is a homogeneous MRDP model version $LC$, with

$$K_2 = 0$$

[1] We do not write it extensively since it involves a rather heavy notation.

$$K_1 = c_i + c_r$$

$$g_1 = c_i + mc_r + c_f$$

$$\delta_1 = \int_0^\infty e^{-\rho\zeta} (\bar{F}(\zeta))^m \sum_{i=1}^m \lambda(\zeta)\ d\zeta$$

and

$$\tilde{\mu}(x) = (m,\ \infty) \qquad \text{for all } x \in X.$$

*Remark 5.1:* Notice that, since the decision variable $d$ is not bounded away from 0, the expected discount factor $a$ is not bounded away from 1. Therefore, the assumptions of model $C$, which are included in the formulation of most other authors ([7], [11], [13], [23], [25]) dealing with MRDP's, are not satisfied.

A replacement policy $\mu$ is a universally measurable function $\mu: S \to A$, and we are looking for an $\epsilon$-optimal policy, where $\epsilon$ is sufficiently small. For this purpose we shall use the algorithm proposed in Section III, with state space reduced to $S$ as discussed in Remark 4.1.

Initially, we set

$$\underline{V} = K_1,\ \bar{V} = g_1/(1-\delta_1),\ \alpha_1 = 0.5$$

and $n_0$ is the smallest integer larger than or equal to $2(\bar{V} - K_1)/K_1$. For any $V$ in $\mathcal{B}_2$ and $(x, l, d)$ in $S \times A$, we have

$$H(V)(x,\ l,\ d) = \int_0^d e^{-\rho\zeta} \prod_{j=1}^m \bar{F}(r_j + \zeta | r_j) \sum_{i=1}^m \lambda(r_i + \zeta)(c_f + V(s_i(\zeta)))\ d\zeta + e^{-\rho d} \prod_{j=1}^m \bar{F}(r_j + d | r_j) V(s_1(d)) + c_i + lc_r$$

where

$$r_j = \begin{cases} x_{j+l-1} & \text{for } j = 1,\ \cdots,\ m-l \\ 0 & \text{for } j = m-l+1,\ \cdots,\ m \end{cases}$$

$$\bar{F}(r_j + \zeta | r_j) = \bar{F}(r_j + \zeta)/\bar{F}(r_j)$$

and $s_i(\zeta)$ is the vector $(r_1 + \zeta, \cdots, r_{i-1} + \zeta, r_{i+1} + \zeta, \cdots, r_m + \zeta)$, which is an element of $S$.

As a numerical illustration, let $m = 3$, $c_i = c_r = 1$, $c_f = 2$, $\rho = 0.1$, and

$$\lambda(t) = 0.02t \text{ for } t \geq 0.$$

This failure rate corresponds to a Weibull distribution. One easily obtains $K_1 = 2$, $g_1 = 6$, $\delta_1 = 0.62011$, $\underline{V} = 2$ and $\bar{V} = 15.88$. We choose $V \equiv 2$ as the initial function, $\alpha_1 = 0.5$, and obtain $n_0 = 14$.

At each iteration of the algorithm (Step 2), let us choose $0 = p_1 < p_2 < \cdots < p_n$ and define $\Omega = \{(p_i, p_j) | 1 \leq j \leq i \leq n\}$. This is the finite set of points at which $T(V)$ is to be evaluated. These points determine a covering of the conical state space $S = \{(x_1, x_2) | x_1 \geq x_2 \geq 0\}$ by $n(n+1)/2$ subsets, as shown in Fig. 1. Among these subsets, $n - 1$ are triangles, $(n-1)(n-2)/2$ are rectangles, and $n$ are unbounded polyedra.

We then choose as follows a functional $V_1$ that interpolates $T(V)$ at the points of $\Omega$: $V_1$ is an interpolating affine function on each triangle, a bilinear function on each bounded rectangle, and an affine function which is constant in $x_1$ on each unbounded polyedron.

Let us take

$$p_i = 2.5(i-1), \qquad i = 1,\ \cdots,\ 5$$

for the first 30 iterations. At the last of these iterations, we obtain $\|(V_1 - V)^+\| = 0.000075$ and $\|(V - V_1)^+\| = 0.0$. Table I gives, for each point $(p_i, p_j)$ in $\Omega$, the value of $T(V)(p_i, p_j)$, as well as the values of $l$ and $d$ for which the minimum is attained in the definition of $T(V)$, after these 30 iterations.

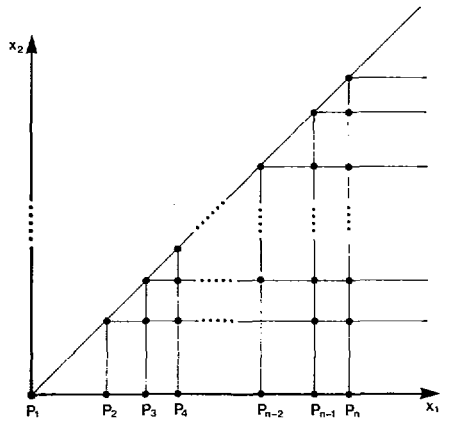



Fig. 1. A partition of the set $X$.

TABLE I
FIRST APPROXIMATION RESULTS, AFTER 30 ITERATIONS. EACH CELL CORRESPONDS TO A POINT IN $\Omega$. THE FIRST NUMBER IN THE CASE IS THE VALUE OF $T(V)$ AT THAT POINT. THE OTHER TWO ARE THE CURRENT BEST VALUES OF $d$ AND $l$

| $p_1$ \ $p_2$ | 0 | 2.5 | 5.0 | 7.5 | 10.0 |
|---|---|---|---|---|---|
| | | | | | 14.022<br>1.45<br>3 |
| | | | | 13.692<br>5.92<br>3 | 13.887<br>3.28<br>3 |
| | | | 13.173<br>7.59<br>3 | 13.455<br>6.76<br>3 | 13.692<br>5.92<br>3 |
| | | 12.371<br>9.26<br>1 | 12.794<br>8.42<br>2 | 13.119<br>7.59<br>2 | 13.388<br>6.76<br>2 |
| | 11.109<br>10.92<br>1 | 11.748<br>10.09<br>1 | 12.199<br>9.26<br>2 | 12.553<br>8.<br>2 | 12.842<br>7.59<br>2 |

Notice that a better approximation of $T(V)(x_1, x_2)$ for $x_1 > 5$ is useless here, since we replace every component whose age is greater than 5. A good approximation to $T(V)$ is useful only in the region where the best value of $l$ is 1.

We then refine the grid, taking

$$p_i = (i-1)/4, \qquad i = 1, \cdots, 22,$$

and do 15 more iterations. At the last iteration, we obtain

$$\|(V_1 - V)^+\| = 0.00008, \quad \|(V - V_1)^+\| = 0.0,$$

and $V_1(0, 0) = 11.148$.

Now, in order to obtain bounds for $V^* - V_1$, we need the values $\delta^-$ and $\delta^+$, which are bounds on $T(V) - V_1$ at the last iteration. Clearly, it is not easy to obtain bounds for $T(V) - V_1$, since this function is defined on a continuous domain. As proposed in Remark 3.1 e) we will thus evaluate $T(V) - V_1$ on a very fine grid, much finer than the preceding one, and estimate $\delta^-$ and $\delta^+$ by the minimum and maximum values of $T(V) - V_1$ on that grid, respectively. This procedure seems reasonable since $V_1$ and $T(V)$ are smooth monotonous functions.

Taking

$$p_i = (i-1)/16, \qquad i = 1, \cdots, 88$$

this yields $\delta^- = 0.00073$ and $\delta^+ = 0.00000$. Using (18), (19), one easily computes the "estimate bounds" for $V^*$

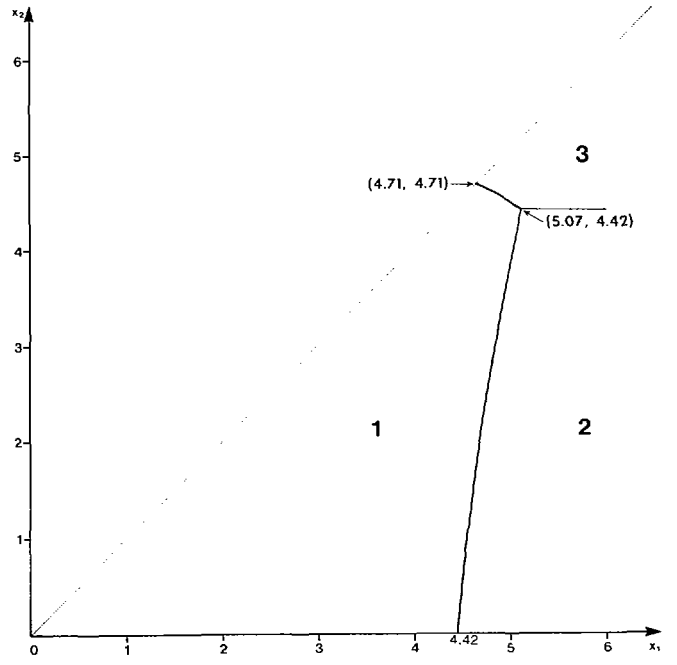



Fig. 2. Number of components to replace as a function of the state of the system, as suggested by the retained policy.

$$-0.021 \le V^* - V_1 \le 0.002.$$

The relative error on $V^*$ is at most 0.2 percent, which is very satisfactory.

In a similar fashion, the optimal value of $d$ can also be interpolated in the same way as $T(V)$, in order to define the finally retained policy $\mu$. We then compute $T_\mu(V) - V_1$ on the finer grid to estimate $\epsilon_0$, and compute the RHS of (23). This expression has a value of 0.021 suggesting that the retained policy is at worst 0.021-optimal.

The conical state space $X$ is partitioned into three regions, according to the number of components that this $\epsilon$-optimal policy tells us to replace. These regions can be seen in Fig. 2. Notice that the number of components to replace is not a monotonously increasing function of the ages of the components. This property was also observed in the discrete-time case (see [10]), and has an easy interpretation. For instance, in state (4.5, 0), replacing two components makes the system completely new, which is profitable. In state (4.5, 4.0), replacing two components leaves one component at age 4.0, while replacing three is more expensive. Replacing only one and waiting for the next intervention happens to be the best action to choose.

## APPENDIX
## PROOFS OF THE THEOREMS OF SECTION III

In the proof of Theorem 3.1, we need the following lemma. This lemma is somewhat related to Proposition 4.6 of Bertsekas and Shreve [5], but different, since the assumption $C$ of [5] is not satisfied here.

*Lemma:* Let $\underline{V} \le \bar{V}$ be two bounded elements of $\mathfrak{B}_1$ and $\mathfrak{B}_3$ be the closed subset of $\mathfrak{B}_1$ defined as $\mathfrak{B}_3 \triangleq \{V \in \mathfrak{B}_1 | \underline{V} \le V \le \bar{V}\}$. If $\varphi: \mathfrak{B}_3 \to \mathfrak{B}_3$ and $\alpha \in [0, 1)$ are such that

$$0 \le \varphi(V_2) - \varphi(V_1) \le \alpha \| V_2 - V_1 \| \qquad \text{(A.1)}$$

for every pair $V_2 \ge V_1$ in $\mathfrak{B}_3$, then there exists $\tilde{V}$ in $\mathfrak{B}_3$ such that for any $V \in \mathfrak{B}_3$, we have

$$\lim_{n \to \infty} \|\varphi^n(V) - \tilde{V}\| = 0 \qquad \text{(A.2)}$$

and

$$-\frac{\alpha}{1-\alpha}\|(V-\varphi(V))^+\| \le \tilde{V}-\varphi(V) \le \frac{\alpha}{1-\alpha}\|(\varphi(V)-V)^+\|. \quad (A.3)$$

*Proof:* First, we show that $\varphi$ is contracting on $\mathfrak{B}_3$. Let $V_1$ and $V_2$ be two arbitrary elements of $\mathfrak{B}_3$. For each $x$ in $X$, let $V_3(x) \triangleq \max(V_1(x), V_2(x))$. $V_3$ is in $\mathfrak{B}_3$, and from (A.1),

$$\varphi(V_1)-\varphi(V_2) \le \varphi(V_3)-\varphi(V_2) \le \alpha\|V_3-V_2\| \le \alpha\|V_1-V_2\|.$$

Since $V_1$ and $V_2$ can be interchanged, we obtain

$$\|\varphi(V_2)-\varphi(V_1)\| \le \alpha\|V_2-V_1\|. \quad (A.4)$$

$\mathfrak{B}_3$ being a closed subset of the Banach space $\mathfrak{B}_0$, the fixed point theorem (see [5, p. 55]) implies that there exists $\tilde{V}$ in $\mathfrak{B}_3$ such that

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

for all $V$ in $\mathfrak{B}_3$. It remains to prove (A.3).

Let $V$ be in $\mathfrak{B}_3$ and for $n = 1, 2, \cdots$, let $\gamma_n$ in $\mathfrak{B}_1$ be defined as

$$\gamma_n(x) = \min\left(\tilde{V}(x)-V(x),\ \|(\varphi(V)-V)^+\|\sum_{i=0}^{n-1}\alpha^i\right).$$

We will show by induction on $n$ that

$$\varphi^{n+1}(V) \le \begin{cases} \varphi(V)+\alpha\|\gamma_n\| \\ V+\gamma_{n+1} \in \mathfrak{B}_3 \end{cases} \quad (A.5)$$

for all $n \ge 2$. Since $\varphi(V)$ is in $\mathfrak{B}_3$, we certainly have $\varphi(V) \le V + \gamma_1$, and since $\gamma_1 \le \tilde{V} - V$, $V + \gamma_1$ is in $\mathfrak{B}_3$.

Assuming that $\varphi^n(V) \le V + \gamma_n \in \mathfrak{B}_3$, we have

$$\begin{aligned}\varphi^{n+1}(V) &= \varphi(\varphi^n(V)) \le \varphi(V+\gamma_n) \\ &\le \varphi(V)+\alpha\|\gamma_n\| \quad (A.6)\\ &\le V+\|(\varphi(V)-V)^+\|+\|(\varphi(V)-V)^+\|\sum_{i=1}^{n}\alpha^i \\ &\le V+\|(\varphi(V)-V)^+\|\sum_{i=0}^{n}\alpha^i. \quad (A.7)\end{aligned}$$

Since $\varphi: \mathfrak{B}_3 \to \mathfrak{B}_3$, we also have

$$\varphi^{n+1}(V) \le \tilde{V}$$

and then

$$\varphi^{n+1}(V) \le V+\gamma_{n+1} \le \tilde{V} \quad (A.8)$$

so $V + \gamma_{n+1}$ is in $\mathfrak{B}_3$. Notice that we have to introduce the functions $\gamma_n$ since we have no guarantee that the RHS in (A.7) belongs to $\mathfrak{B}_3$, and so we cannot directly apply $\varphi$ on this function. Equation (A.5) now follows from (A.6) and (A.8). Taking the limit in (A.6), we obtain

$$\begin{aligned}\tilde{V} &= \lim_{n\to\infty}\varphi^{n+1}(V) \le \varphi(V)+\alpha\lim_{n\to\infty}\|\gamma_n\| \\ &\le \varphi(V)+\|(\varphi(V)-V)^+\|\alpha/(1-\alpha).\end{aligned}$$

In a similar way, we can show that

$$\tilde{V} \ge \varphi(V)-\frac{\alpha}{1-\alpha}\|(V-\varphi(V))^+\|$$

and that completes the proof. ■

*Proof of Theorem 3.1:* a) First, we show by induction on $n$ that for $n = 1, \cdots, n_0$, we have

$$-n\delta^- -(n-1)\|(V-V_1)^+\| \le T^n(V)-V_1 \le n\delta^+ +(n-1)\|(V_1-V)^+\|.$$

For $n = 1$, it follows directly from the definitions. Assume that it is true for $n - 1$, where $1 \le n - 1 < n_0$. Then

$$\begin{aligned}T^n(V)-V_1 &= T^n(V)-T^{n-1}(V)+T^{n-1}(V)-V_1 \\ &\le \|(T(V)-V)^+\|+(n-1)\delta^+ +(n-2)\|(V_1-V)^+\| \\ &\le \|(T(V)-V_1)^+\|+\|(V_1-V)^+\| \\ &\quad +(n-1)\delta^+ +(n-2)\|(V_1-V)^+\| \\ &\le n\delta^+ +(n-1)\|(V_1-V)^+\|\end{aligned}$$

and in a similar way,

$$T^n(V)-V_1 \ge -n\delta^- -(n-1)\|(V-V_1)^+\|.$$

Letting $n = n_0$, we thus obtain

$$-\epsilon^- \le T^{n_0}(V)-V_1 \le \epsilon^+. \quad (A.9)$$

On the other hand, if $V_2 \ge V_1$ are two functions in $\mathfrak{B}_2$, then by [15, Lemmas 3 and 7], we have

$$0 \le T^{n_0}(V_2)-T^{n_0}(V_1) \le \alpha_1\|V_2-V_1\|. \quad (A.10)$$

Applying the preceding lemma (with $\mathfrak{B}_3 = \mathfrak{B}_2$, $\varphi = T^{n_0}$ and $\tilde{V} = V^*$), we obtain

$$V^* \begin{cases} \ge T^{n_0}(V)-\dfrac{\alpha_1}{1-\alpha_1}\|(V-T^{n_0}(V))^+\| \\ \le T^{n_0}(V)+\dfrac{\alpha_1}{1-\alpha_1}\|(T^{n_0}(V)-V)^+\|. \end{cases} \quad (A.11)$$

Using (A.9) and (A.11), (18), (19) follow easily.

b) The first statement follows from [5, Proposition 7.50]. In (22), the first inequality is obvious by definition of $V^*$, and it remains to prove the second inequality. From [5, Lemma 7.30 and Proposition 7.48] we see that $T_\mu(W) \in \mathfrak{B}_1$ for each $W$ in $\mathfrak{B}_1$. For the version $C$ of the model, by [15, Lemma 3] we also have $T_\mu: \mathfrak{B}_2 \to \mathfrak{B}_2$.

For the version $LC$, define $\underline{V} \equiv K_1 + \min(0, K_2)$, $g_2 \equiv g_1/(1-\delta_1)+\epsilon_0-\delta^+ + \|\underline{V}\|$, $\tilde{V} \equiv (g_2+(1+\alpha)\|\underline{V}\|)/(1-\alpha)$ and $\mathfrak{B}_3$ as in the previous lemma. For each $W$ in $\mathfrak{B}_3$, we have $T_\mu(W) \ge T_\mu(\underline{V}) \ge T(\underline{V}) \ge \underline{V}$. From (13), (14) and since $V$ is in $\mathfrak{B}_2$, we have

$$T_\mu(V)(x) \ge g(x, \mu(x))-\|\underline{V}\|, \quad \text{for all } x \text{ in } X. \quad (A.12)$$

From (20), (A.12) and since $V_1$ is in $\mathfrak{B}_2$, we obtain

$$g(x, \mu(x)) \le g_1/(1-\delta_1)+\epsilon_0-\delta^+ +\|\underline{V}\| = g_2. \quad (A.13)$$

From (21), (12)–(14) and (A.13), we thus have

$$\begin{aligned}T_\mu(W) &= T_\mu(\underline{V})+T_\mu(W)-T_\mu(\underline{V}) \\ &\le g_2+\|\underline{V}\|+\alpha\|W-\underline{V}\| \\ &\le g_2+\|\underline{V}\|+\alpha(\|\tilde{V}\|+\|\underline{V}\|) \\ &\le (g_2+(1+\alpha)\|\underline{V}\|)(1+\alpha/(1-\alpha)) = \tilde{V}.\end{aligned}$$

Thus, $T_\mu: \mathfrak{B}_3 \to \mathfrak{B}_3$. For both versions of the model, the previous lemma applies (with $\mathfrak{B}_3 = \mathfrak{B}_2$ for version $C$, $\mathfrak{B}_3$ as defined above for version $LC$, $\varphi = T_\mu$ and $\tilde{V} = \tilde{V}_\mu$). Further, as in the proof of Lemma 10 in [15], we can show that $\tilde{V}_\mu = V_\mu$. We then obtain, since $V \in \mathfrak{B}_2 \subseteq \mathfrak{B}_3$,

$$V_\mu \le T_\mu(V)+\frac{\alpha}{1-\alpha}\|(T_\mu(V)-V)^+\|. \quad (A.14)$$

Using (20) in (A.14), we obtain

$$V_\mu \le V_1+\epsilon_0+\frac{\alpha}{1-\alpha}\|(V_1+\epsilon_0-V)^+\|$$

and from (18), (22) follows. ■

*Proof of Theorem 3.3.* a) Let $\epsilon > 0$. Choose a positive integer $i$ such that

$$(\alpha_1)^i < \frac{\epsilon}{2\|\bar{V} - \underline{V}\|}$$

where $\underline{V}$ and $\bar{V}$ are the bounds of $\mathfrak{B}_2$, and then $k > i$ such that

$$\max\ (\delta_j^-,\ \delta_j^+) < \frac{(1-\alpha_1)\epsilon}{2n_0}$$

for all $j \geq n_0(k-i)$. For each integer $n > n_0 k$, we obtain using [15, Lemmas 3 and 7] and equations (12)–(15)

$$\begin{aligned}
\|V_n - V^*\| &\leq \|V^* - T^{in_0}(V_{n-in_0})\| + \|T^{in_0}(V_{n-in_0}) - V_n\| \\
&\leq \|T^{in_0}(V^*) - T^{in_0}(V_{n-in_0})\| + \sum_{j=1}^{i} \|T^{jn_0}(V_{n-jn_0}) \\
&\quad - T^{(j-1)n_0}(V_{n-(j-1)n_0})\| \\
&\leq (\alpha_1)^i \|V^* - V_{n-in_0}\| \\
&\quad + \sum_{j=1}^{i} (\alpha_1)^{j-1} \|T^{n_0}(V_{n-jn_0}) - V_{n-(j-1)n_0}\| \\
&\leq (\alpha_1)^i \|\bar{V} - \underline{V}\| + \sum_{j=1}^{i} (\alpha_1)^{j-1} \sum_{r=1}^{n_0} \|T^r(V_{n-(j-1)n_0-r}) \\
&\quad - T^{r-1}(V_{n-(j-1)n_0-r+1})\| \\
&\leq (\alpha_1)^i \|\bar{V} - \underline{V}\| + \sum_{j=1}^{i} (\alpha_1)^{j-1} \sum_{r=1}^{n_0} \|T(V_{n-(j-1)n_0-r}) \\
&\quad - V_{n-(j-1)n_0-r+1}\| \\
&< \epsilon/2 + \frac{1}{1-\alpha_1} n_0 \frac{(1-\alpha_1)\epsilon}{2n_0} = \epsilon.
\end{aligned}$$

Since $\epsilon$ is arbitrary, this completes the proof.

b) From a), the sequence of values of $\|V - V^*\|$ obtained in the algorithm converges to 0, as well as the sequence of values of $\|T(V) - V\|$, since $\|T(V) - V\| < \|T(V) - V^*\| + \|V^* - V\| < 2\|V - V^*\|$. The sequence of values of $\|V_1 - V\|$ also converges to 0, since

$$\begin{aligned}
\|V_1 - V\| &\leq \|T(V) - V\| + \|T(V) - V_1\| \\
&\leq \|T(V) - V\| + \max\ (\delta^-,\ \delta^+).
\end{aligned}$$

Therefore, since $\|(V - V_1)^+\|$ and $\|(V_1 - V)^+\|$ are bounded by $\|V_1 - V\|$, the values of $\epsilon^-$ and $\epsilon^+$ also converge to 0. Taking $\alpha = \alpha_1$, the left-hand expression in (23) is smaller or equal to

$$\epsilon^- + \frac{\alpha_1}{1-\alpha_1}(\|(V - V_1)^+\| + \epsilon^- + \|(V_1 - V)^-\| + \epsilon_0/\alpha_1).$$

This expression tends to 0, and will thus become smaller than $\epsilon$ after a finite number of iterations. ■

## Acknowledgment

The authors wish to thank the associate editor and one of his students for their careful reading and comments which helped to correct and improve the final version of this paper. All possible remaining errors are the sole responsibility of the authors.

## References

[1] A. Alj and A. Haurie, "Hierarchical control of a population process with application to group preventive maintenance," in *Proc. IFAC 2nd Symp. on Large Scale Systems Theory and Appl.*, Toulouse, France, June 1980, pp. 24–26.

[2] R. Bellman and S. E. Dreyfus, *Applied Dynamic Programming*. Princeton, NJ: 1962.

[3] C. Berger, "Commande optimale de systèmes stochastiques markoviens de grande dimension," Ecole des H.E.C., Montréal, les cahiers du GERAD, G-80-04, 1980.

[4] D. P. Bertsekas, "Convergence of discretization procedures in dynamic programming," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 415–419, 1975.

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

[6] J. W. Daniel, "Splines and efficiency in dynamic programming," *J. Math. Anal. Appl.*, vol. 54, pp. 402–407, 1976.

[7] E. V. Denardo, "Contraction mappings in the theory underlying dynamic programming," *SIAM Rev.*, vol. 9, pp. 165–177, 1967.

[8] B. L. Fox, "Discretizing dynamic programs," *J. Optimiz. Theory Appl.*, vol. II, pp. 228–234, 1973.

[9] I. B. Gertsbakh, *Models of Preventive Maintenance*. Amsterdam, The Netherlands: North-Holland, 1977.

[10] A. Haurie and P. L'Ecuyer, "A stochastic control approach to group preventive replacement in a multicomponent system," *IEEE Trans. Automat. Contr.*, vol. AC-27, no. 2, pp. 387–393, 1982.

[11] K. Hinderer, "Foundations of non-stationary dynamic programming with discrete time parameter," (Lecture Notes in Oper. Res. and Math. Syst.). New York: Springer-Verlag, 1970.

[12] ——, "On approximate solutions of finite stage dynamic programs," in *Dynamic Programming and its Applications*, M. L. Puterman, Ed. New York: Academic, 1978, pp. 289–317.

[13] H. J. Langen, "Convergence of dynamic programming models," *Math. Operat. Res.*, vol. 6, no. 4, pp. 493–512, 1981.

[14] P. L'Ecuyer, "Processus de décision markoviens à étapes discrètes: Application à des problèmes de remplacement d'équipement", Ph.D. dissertation, Dep. Inform. R.O., Univ. Montréal, Apr. 1983; also Ecole des H.E.C., Montréal, cahiers du GERAD G-83-06, 1983.

[15] P. L'Ecuyer and A. Haurie, "Discrete event dynamic programming in Borel spaces with state dependent discounting," Dep. Inform., U. Laval, Québec, Rep. DIUL-RR-8309, 1983.

[16] J. MacQueen, "A modified dynamic programming method for Markovian decision problems," *J. Math. Anal. Appl.*, vol. 14, pp. 38–43, 1966.

[17] J. MacQueen, "A test for suboptimal actions in Markovian decision problems," *Oper. Res.*, vol. 15, pp. 559–561, 1967.

[18] T. L. Morin, "Computational advances in dynamic programming," in *Dynamic Programming and its Applications*, M.L. Puterman Ed. New York: Academic, 1978, pp. 53–90.

[19] E. Porteus, "Some bounds for discounted sequential decision processes," *Management Sci.*, vol. 18, pp. 7–11, 1971.

[20] ——, "Bounds and transformations for discounted finite markov decision chains," *Oper. Res.*, vol. 23, pp. 761–784, 1975.

[21] ——, "On the optimality of structured policies in countable stage decision processes," *Management Sci.*, vol. 22, pp. 148–157, 1975.

[22] M. L. Puterman and M. C. Shin, "Action elimination procedures for modified policy iteration algorithms," *Oper. Res.*, vol. 30, no. 2, pp. 301–318, 1982.

[23] S. Ross, *Apply Probability Models with Optimization Applications*. San Francisco, CA: Holden Day, 1970.

[24] R. Rishel, "Group preventive maintenance: An example of controlled jump processes," in *Proc. 20th IEEE Conf. Decision Contr.*, San Diego, CA, Dec. 1981, pp. 786–791.

[25] M. Schäl, "Conditions for optimality in dynamic programming and for the limit of n-stages optimal policies to be optimal," *Z. Warschein. verw. Gebiete*, vol. 32, pp. 179–198, 1975.

[26] L. C. Thomas, R. Arley, and A. C. Lavercombe, "Computational comparison of value iteration algorithms for discounted Markov decision process," *Oper. Res. Lett.*, vol. 2, pp. 72–76, June 1983.

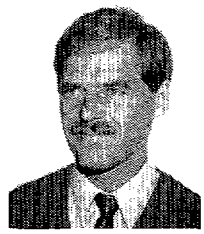
[27] D. J. White, "Elimination of non optimal actions in Markov decision process," *in Dynamic Programming and Its Applications*, M. L. Puterman Ed. New York: Academic, 1979.

[28] W. Whitt, "Approximation of dynamic programs I and II," *Math. Oper. Res.*, vol. 3, pp. 231–243, 1978; also in vol. 4, pp. 179–185, 1979.

[29] P. Whittle, "A simple condition for regularity in negative programming," *J. Appl. Prob.*, vol. 16, pp. 305–318, 1979.

**Alain Haurie** (M'74) was born in Algiers, Algeria, on August 26, 1940. He received the Licence es Sciences degree in mathematics from the University of Algiers, Algeria, in 1961, the Doctorat de 3e cycle degree in applied mathematics from the University of Paris 7, Paris, France, in 1970, and the Doctorat es Sciences (Doctorat d'état) degree also from the University of Paris 7.

Since 1963 he has been a Professor at l'Ecole des Hautes Etudes Commerciales de Montréal, which is the Graduate Business School of the University of Montréal, Montréal, P.Q., Canada. In 1976 and 1977 he was on leave of absence at INSEA, Rabat, Morocco. From 1970 to 1973 he held a part-time teaching and research position in the Department of Mathematics of l'Ecole Polytechnique de Montréal where he was responsible for a graduate course on Optimal Control Theory. In 1979, he held a similar position in the Department of Operation Research, University of Montréal. Since 1980 he has been Director of GERAD (Groupe d'études et de recherche en analyse des décisions). His current research interests includ· ıpplication of stochastic control theory to societal problems, application of optimal control theory to economic planning, modeling of manufacturing systems, and game theory.



**Pierre L'Ecuyer** (M'83) was born in Rimouski, Canada, in 1950. He received the B.Sc. degree in mathematics in 1972, and the M.Sc. and Ph.D. degrees in operations research, in 1980 and 1983, respectively, both from the University of Montréal, Montréal, P.Q., Canada.

He taught college mathematics from 1973 to 1978. From 1980 to 1983, he was a Research Assistant at l'Ecole des Hautes Etudes Commerciales, Montréal, P.Q., Canada. He is presently an Adjoint Professor in Computer Science at Laval University, Ste-Foy, Québec, P.Q., Canada. His research interests are in Markov renewal decision processes, approximation methods in dynamic programming, discrete-event simulation, and software engineering.

Dr. L'Ecuyer is a member of the Operations Research Society of America, the SCS, and the Association for Computing Machinery.