ISSN 0711-2440

APPROXIMATION AND BOUNDS IN DISCRETE EVENT DYNAMIC PROGRAMMING

by

A. Haurie and P. L'Ecuyer

G-83-25

Octobre 1983

Revisé Juin 1984

Résumé: Cet article concerne un algorithme général de programmation dynamique adapté à la résolution de problèmes de commande optimale de systèmes stochastiques à événements discrets.

Un intérêt particulier est porté aux techniques numériques utilisées pour résoudre par approximations successives l'équation de la programmation dynamique.

L'algorithme et les méthodes numériques sont illustrés à partir d'un problème d'entretien préventif d'un équipement complexe.

APPROXIMATION AND BOUNDS IN DISCRETE EVENT DYNAMIC PROGRAMMING*

A. Haurie** P. L'Ecuyer***

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.

---

* Research supported by SSHRC-Canada Grant #410-81-0722, NSERC-Canada Grant #A4952, FCAC-Québec Grant #EQ-0428 to the first author and NSERC-Canada Grant #A5463 to the second author.

** GERAD, Ecole des Hautes Etudes Commerciales, 5255 avenue Decelles, Montréal (Québec) Canada H3T 1V6, Tel. 514-343-4443.

*** Département d'informatique, Université Laval, Cité Universitaire, Québec, Canada G1K 7P4, Tel. 418-656-3226.

1. 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 queuing 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, 23] 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 [25] obtained a condition (the bridging condition) for a Markov decision process with non negative costs to enjoy regularity properties inplying the limit of n-state optimal policies to be optimal.

Finally a generalization of the Bertsekas Shreve dynamic programming formalism has been achieved in Refs. [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 Ref. [10], in the setting of an optimal group replacement problem.

A more satisfactory approach consists in 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 2 we review briefly the most relevant results of [15] which concern a gneeral multistage decision process; in section 3 we present the use of approximation techniques for the solution of the dynamic programming equation; in section 4, we see that our model encompasses the Markov-renewal decision processes with borelian state and action spaces; finally in section 5 we use the approach for the numerical solution of a complex maintenance problem.

2. 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 are given in Refs. [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 Γ of $X \times A$, called the set of admissible state-action couples, such that

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

The dynamics of the system is described by a stochastic kernel Q, which is a family $\{Q(\cdot|x,a)\ ,\ (x,a) \in X \times A\}$ 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 $g: \Gamma \to \mathbb{R}$. 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 $Q(\cdot|x_n, a_n)$. The cost incurred at stage n is discounted to an arbitrary 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).

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

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

Associated with any initial state x and control law, there is a uniquely defined probability measure $P_{\mu,x}$ (see Refs. [14], [15]) on the set H of infinite sequences $(x_0, a_0, x_1, a_1, \ldots)$ where $(x_i, a_i) \in X \times A$ for $i = 0, 1, \ldots$ 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

$$\int_H (\sum_{i=0}^{\infty} c_i) d P_{\mu,x} \tag{4}$$

is also well defined.

We introduce, when they exist, the values

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

and

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

where U is the set of admissible control laws.

A control law μ in U is said to be optimal [resp. ε-optimal] if $V_\mu(x) = V^*(x)$ [resp. $V_\mu(x) \leq V^*(x) + \varepsilon$] for all x in X.

Remark 2.1: 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 analysed in Ref. [5] stems from the consideration of a discount factor defined as a function $\beta(x_n)$ instead of β^n, with $\beta \in (0,1)$.

In the latter case, the condition that β 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, α and Q.

Assumption 1.1: There exists α_1, g_0, g_1 such that

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

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 numbers δ_1, K_1, K_2, g_1 in $\mathbb{R}_+$ such that

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

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

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

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

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

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

$$\mathcal{B}_2 = \begin{cases} \{V \in \mathcal{B}_1 :\ \dfrac{g_o}{1-\alpha_1} \le V \le \dfrac{g_1}{1-\alpha_1}\} & \text{for version C} \\ \\ \{V \in \mathcal{B}_1 :\ K_1 + \min(0,K_2) \le V \le \dfrac{g_1}{1-\alpha_1}\} & \text{for version LC.} \end{cases}$$

For any V in $\mathcal{B}_2$, we introduce

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

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

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

The following two theorems summarize the dynamic programming approach.

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

(a) $T(V) = V \quad \text{iff} \quad V = V^*$

(b) $\text{If} \quad T(V) \leq V \quad \text{then} \quad V^* \leq V$

(c) $\text{If} \quad T(V) \geq V \quad \text{then} \quad V^* \geq V$

(d) $\lim_{n\to\infty} \|T^n(V) - V^*\| = 0$

(e) $V^* \in \mathcal{B}_2$

(f) A control law μ is optimal iff $T_\mu(V^*) = V^*$

(g) A control law μ is optimal iff $T(V_\mu) = V_\mu \in \mathcal{B}_2$.

Remark 2.2: 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 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 μ iff the $\inf_{a\in A(x)} H(V^*)(x)$ is attained for all x in X.

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

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

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

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

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

Remark 2.3: 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, g.β and V are lower semi-continuous and Q is continuous X×A.

Furthermore (b) is a constructive proof of the existence of an optimal control law.

3. APPROXIMATION TECHNIQUES FOR THE SOLUTION OF THE D.P. EQUATION

One of the last "frontier" in the theory and applications of Dynamic Programming lies in the numerical solution of the D.P. equation by using approximation techniques. For example Rishel [22] 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 D.P. equation given in section 2, and generalizes most of the approaches proposed in the O.R. literature for the computation of approximations and bounds in dynamic programs.

The value iteration technique consists in applying the D.P. operator T repeatedly, until some convergence test has been passed. McQueen [17], Denardo [7], 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 in 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 [24] 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_μ is the value-function obtained on X when one uses the optimal policy μ 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 $\mathcal{B}_1$ we denote

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

Theorem 3.1: Let α_1 satisfying (6) and $n_0 = 1$ for version C. Let $\alpha_1 \in (0,1)$ and

$$n_0 \triangleq \frac{\frac{g_1}{1-\delta_1} - K_1 - \min(0, K_2)}{\alpha_1(K_1 + K_2)}$$

for version LC of the model.

Consider two functions V and V_1 in $\mathcal{B}_2$, and two numbers δ^-, δ^+ in $\mathbb{R}_+$ such that

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

Define

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

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

Then the following holds:

a)
$$V^* \begin{cases} \geq V_1 - \varepsilon^- - \dfrac{\alpha_1}{1-\alpha_1}\,\|(V-V_1+\varepsilon^-)^+\| & (18) \\ \leq V_1 + \varepsilon^+ + \dfrac{\alpha_1}{1-\alpha_1}\,\|(V_1+\varepsilon^+-V)^+\| & (19) \end{cases}$$

b) For any $\varepsilon_0 > 0$, there exists a control law μ such that

$$T_\mu(V) \leq V_1 + \varepsilon_0. \tag{20}$$

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 $(V_2,V_3) \in \mathcal{B}_1$ satisfying $V_3 \geq V_2$, then

$$V^* \leq V_\mu \leq V^* + \varepsilon^- + \frac{\alpha_1}{1-\alpha_1}\,\|(V-V_1+\varepsilon^-)^+\| + \varepsilon_0 + \frac{\alpha}{1-\alpha}\,\|(V_1+\varepsilon_0-V)^+\| \quad \blacksquare \tag{22}$$

This theorem is complemented by the following one, due originally to McQueen [1], which permits the elimination of non optimal actions.

Theorem 3.2: If $\underline{V}$ and $\bar{V}$ are two functions in $\mathcal{B}_2$ 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:

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

These two theorems suggest the following algorithm:

ALGORITHM: (1) Set α_1 and n_o as in Theorem 3.1.

Set $\tilde{\mathcal{B}} := \mathcal{B}_2$.

Choose an upper bound for the total number of iterations.

Choose $\varepsilon > 0$.

Take any function V in $\tilde{\mathcal{B}}$ as an initial guess of V^*.

(2) Compute T(V) at a finite number of points in X.

Define V_1 in $\tilde{\mathcal{B}}$ as an approximation of T(V) on X.

(3) Compute δ^-, δ^+ in $\mathbb{R}_+$ such that

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

and compute ε^- and ε^+ 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 ε.

(4) Find a control law μ and $\varepsilon_0 \geq 0$ such that

$$T_\mu(V) \leq V_1 + \varepsilon_0 \quad .$$

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

$$T_\mu(V_3) - T_\mu(V_2) \leq \alpha \, \|V_3 - V_2\|$$

for any pair $(V_2, V_3) \in (\mathcal{B}_0 \cap \mathcal{B}_1)^2$ and if

$$\varepsilon^- + \frac{\alpha_1}{1-\alpha_1} \|V - V_1 + \varepsilon^-)^+\| + \varepsilon_0 + \frac{\alpha}{1-\alpha} \|V_1 + \varepsilon_0 - V)^+\| \leq \varepsilon \qquad (23)$$

then μ is a control law which is ε-optimal.

Stopping rule 2: Stop when one has obtained an ε-optimal control law.

(5) Stopping rule 3: Stop when the maximal number of iterations has been attained.

(6) Set

$$\underline{V} := V_1 - \varepsilon^- - \frac{\alpha_1}{1-\alpha_1} \|(V - V_1 + \varepsilon^-)^+\|$$

$$\bar{V} := V_1 + \varepsilon^+ + \frac{\alpha_1}{1-\alpha_1} \|(V_1 + \varepsilon^+ - V)^+\|$$

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

$$V := V_1$$

(7) Go to (2) .

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

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

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

be such that

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

with

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

Then

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

(b) For the version C of the model, if the sequences of δ^-, δ^+, and ε_0 values obtained in steps (3) and (4) of the algorithm converge to 0, then for any $\varepsilon > 0$ an ε-optimal control law is obtained in a finite number of iterations.

Remarks 3.1: (a) After step (6) of the algorithm, an elimination of non optimal actions can be done by using Theorem 3.2. Also the parameters α_1 and n_0 could be revaluated.

(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 d.p. procedure, and the errors due to the fact that only a finite number of iterations are effectuated.

(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 δ^-, δ^+ and ε_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) can be computed ponctually 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 $a = x_1 < x_2 < \ldots < x_n = b$, defines

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

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 δ^- and δ^+, 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 choosen, numerous, and if T(V)-V behaves reasonably, then the smallest and largest of these errors can be taken as estimates of δ^- and δ^+ respectively.

4. 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, have a state set X defined as a cartesian product

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

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 occurence 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 ρ 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 2.

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 (ζ, 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 ζ is non zero. The cost of transition from stage n to stage n+1 is given by

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

We might be interested in the subclass of homogeneous Markov-renewal decision processes 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 μ 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

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

Further, if A(t,s), g(t,s,a) and $Q(\cdot \mid t,s,a)$ are independant of t, then

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

If V(t,s) is independent of t, then H(V)(t,s,a) and T(V)(t,s) are also independant of t. Hence, in this case, the state space can be reduced to S. We use this simplification in the following example.

5. EXAMPLE: A MULTICOMPONENT SYSTEM WITH IDENTICAL ELEMENTS

Consider a system comprised of m identical and stochastically independent components, each having a known and non decreasing 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 [22], 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 $\ell \in\{1,\ldots,m\}$ 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 ∞.

Owing to the non-decreasing 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

$$X = \{(x_1,x_2,\ldots,x_{m-1}) \in \mathbb{R}^{m-1} \mid x_1 \geq x_2 \geq \ldots \geq x_{m-1} \geq 0\}$$

and

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

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

$$K_2 = 0$$

$$K_1 = c_i + c_r$$

$$g_1 = c_i + m\, c_r + c_f$$

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

and

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

A replacement policy μ is a universally measurable function $\mu : X \to A$, and we are looking for an ε-optimal policy, where ε is sufficiently small. For this purpose we shall use the algorithm proposed in section 3.

Initially, we set

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

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 Γ, we have

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

$$+ e^{-\rho d} \prod_{j=1}^{m} \bar{F}(r_j+d \mid r_j)\; V(s_1(d)) + c_i + \ell\, c_r$$

---

* We do not write it extensively since it involves a rather heavy notation.

where

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

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

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

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

$$\lambda(t) = .02t \qquad \text{for} \qquad t \geq 0.$$

This failure rate corresponds to a Weibull distribution. One easily obtains $K_1 = 2$, $g_1 = 6$, $\delta_1 = .62011$, $\underline{V} = 2$ 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 $0 = p_1 < p_2 < \ldots < p_n$ and define $\Omega = \{(p_i,p_j) \mid 1 \leq j \leq i \leq n\}$. 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 $X = \{(x_1,x_2) \mid x_1 \geq x_2 \geq 0\}$ by $n(n+1)/2$ subsets, as shown in figure 5.1. Among these subsets, n-1 are triangles, $(n-1)(n-2)/2$ are rectangles and n are unbounded polyedra.

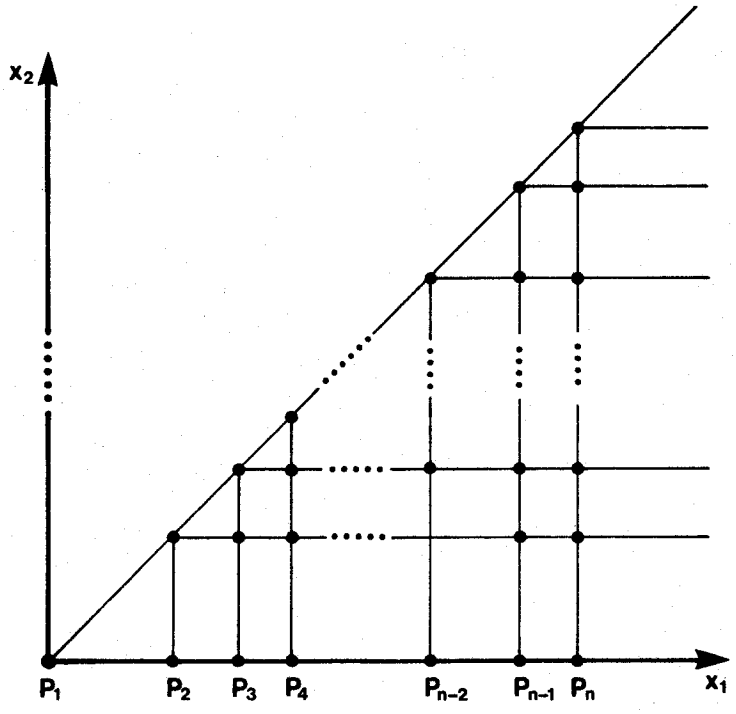

Figure 5.1: A partition of the set X.

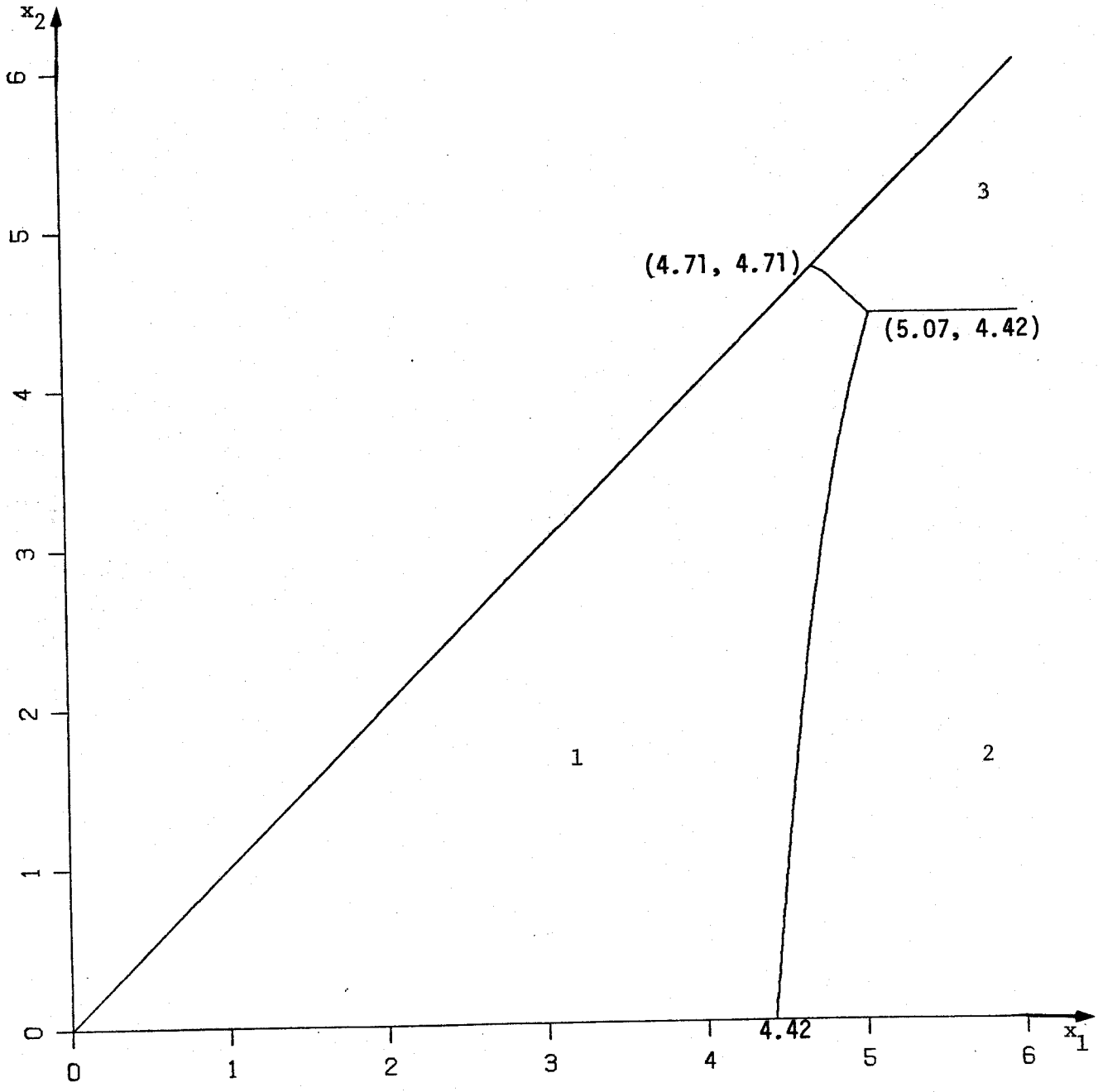

Figure 5.3: Number of components to replace as a function of the state of the system, as suggested by the retained policy.

We then choose as follows a functional V_1 that interpolates T(V) at the points of Ω : 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

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

for the first 30 iterations. At the last of these iterations, we obtain $\|(V_1-V)^+\| = .000075$ and $\|(V-V_1)^+\| = 0.0$. Table 5.2 gives, for each point (p_i,p_j) in Ω, 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.

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

Table 5.2: First approximation results, after 30 iterations. Each case corresponds to a point in Ω. 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 respectively.

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

$$p_i = (i-1)/4 \quad , \quad i = 1, \ldots, 22,$$

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

$$\|(V_1-V)^+\| = .00008 \quad , \quad \|(V-V_1)^+\| = 0.0 \ , \text{ and } \ V_1(0,0) = 11.148.$$

Now, in order to obtain bounds for V^*-V_1, we need the values δ^- and δ^+, 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 δ^- and δ^+ 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

$$p_i = (i-1)/16 \quad , \quad i = 1,\ldots,88 \quad ,$$

this yields $\delta^- = .00073$ and $\delta^+ = .00000$. Using Eqs (18, 19), one easily computes the "estimate bounds" for V^*

$$-.021 \leq V^* - V_1 \leq .002 \quad .$$

The relative error on V^* is at most 0.2%, 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 μ. We then compute $T_\mu(V)-V_1$ on the finer grid to estimate ε_0, and compute the RHS of Eq. (23). This expression has a value of .021 suggesting that the retained policy is at worst .021-optimal.

The conical state space X is partitioned into three regions, according to the number of components that this ε-optimal policy tells us to replace. These regions can be seen in Figure 5.3. 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 3

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

$$0 \leq \varphi(V_2) - \varphi(V_1) \leq \alpha \|V_2 - V_1\| \tag{A.1}$$

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

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

and

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

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

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

Since V_1 and V_2 can be interchanged, we obtain

$$\|\varphi(V_2) - \varphi(V_1)\| \leq \alpha \|V_2 - V_1\|. \tag{A.4}$$

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

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

for all V in $\mathcal{B}_3$.

Let V be in $\mathcal{B}_3$ and γ_1 in $\mathcal{B}_1$ defined as

$$\gamma_1(x) \triangleq \min\ (\bar{V}(x) - V(x),\ \|(\varphi(V)-V)^+\|) \geq 0.$$

We have $\varphi(V)-V \leq \bar{V} - V$ since $\varphi(V)$ is in $\mathcal{B}_3$, and also $\varphi(V)-V \leq \|(\varphi(V)-V)^+\|$. Hence $\varphi(V)-V \leq \gamma_1$ and since $\gamma_1 \leq \bar{V}-V$, we obtain

$$\varphi(V) \leq V + \gamma_1 \quad \in \quad \mathcal{B}_3. \tag{A.5}$$

Notice that we had to introduce γ_1 because we do not know if $V + \|(\varphi(V)-V)^+\|$ is in $\mathcal{B}_3$.

Using (A.5) and (A.1), we obtain

$$\varphi^2(V) \leq \varphi(V + \gamma_1) \leq \varphi(V) + \alpha\|\gamma_1\|$$

$$\leq V + (1 + \alpha)\ \|\gamma_1\| \ .$$

This can be repeated, to show by induction that

$$\varphi^{n+1}(V) \leq \varphi(V) + \|\gamma_1\| \sum_{i=1}^{n} \alpha^i \leq V + \|\gamma_1\| \sum_{i=0}^{n} \alpha^i \ .$$

Taking the limit, we obtain

$$\tilde{V} = \lim_{n\to\infty} \varphi^{n+1}(V) \leq \varphi(V) + \frac{\alpha}{1-\alpha}\|(\varphi(V)-V)^+\| \ .$$

In a similar way, we can show that

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

and that completes the proof. ■

Proof of theorem 3.1. (a) First, we show that

$$-\varepsilon^- \leq T^{n_0}(V)-V_1 \leq \varepsilon^+ \ . \tag{A.6}$$

If $n_0 = 1$, (A.6) follows directly from the definitions. Assume that it is true for all $n_0 \leq n-1$, where $n \geq 2$ is fixed. Then,

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

and in a similar way,

$$T^n(V)-V_1 \leq -\,\varepsilon^- \,.$$

It follows by induction that (A.6) holds.

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

$$0 \leq T^{n_o}(V_2) - T^{n_o}(V_1) \leq \alpha_1\,\|V_2-V_1\| \,. \tag{A.7}$$

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

$$V^* \begin{cases} \geq T^{n_o}(V) - \dfrac{\alpha_1}{1-\alpha_1}\,\|(V-T^{n_o}(V))^+\| \\ \leq T^{n_o}(V) - \dfrac{\alpha_1}{1-\alpha_1}\|(T^{n_o}(V)-V)^+\| \,. \end{cases} \tag{A.8}$$

Using (A.6) and (A.8), (18-19) follow easily.

(b) The first statement follows from proposition 7.50 in [5]. In (22), the first inequality is obvious by definition of V^*, and it remains to prove the second inequality. From lemma 7.30 and proposition 7.48 in [5], we see that $T_\mu(V) \in \mathcal{B}_1$ for each V in $\mathcal{B}_1$. For the version C of the model, by lemma 3 in [15], we also have $T_\mu: \mathcal{B}_2 \to \mathcal{B}_2$. For the version LC, define $\underline{V} \equiv K_1 + \min(0, K_2)$, $\overline{V} \equiv (g_1 + (1+\alpha)\|\underline{V}\|)/(1-\alpha)$ and $\mathcal{B}_3$ as in the previous lemma. For each V in $\mathcal{B}_3$, we have $T_\mu(V) \geq T_\mu(\underline{V}) \geq T(\underline{V}) \geq \underline{V}$, and using (21), (13,14) and (11,12),

$$\begin{aligned} T_\mu(V) &= T_\mu(\underline{V}) + T_\mu(V) - T_\mu(\underline{V}) \\ &\leq g_1 + \|\underline{V}\| + \alpha\|V-\underline{V}\| \\ &\leq g_1 + \|\underline{V}\| + \alpha(\|\overline{V}\| + \|\underline{V}\|) \\ &\leq (g_1 + (1+\alpha)\|\underline{V}\|)(1 + \alpha/(1-\alpha)) = \overline{V} . \end{aligned}$$

Thus, $T_\mu: \mathcal{B}_3 \to \mathcal{B}_3$. For both versions of the model, the previous lemma applies (with $\mathcal{B}_3 = \mathcal{B}_2$ for version C, $\mathcal{B}_3$ 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 $V \in \mathcal{B}_2 \subseteq \mathcal{B}_3$,

$$V_\mu \leq T_\mu(V) + \frac{\alpha}{1-\alpha}\|(T_\mu(V)-V)^+\| . \qquad (A.9)$$

Using (20) in (A.9), we obtain

$$V_\mu \leq V_1 + \varepsilon_0 + \frac{\alpha}{1-\alpha}\|(V_1 + \varepsilon_0 - V)^+\|$$

and from (18-19), (22) follows. ■

Proof of theorem 3.3. (a) Let $\varepsilon > 0$. Choose a positive integer i such that

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

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

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

for all $j \geq n_0(k-i)$. For each integer $n > n_0 k$, we obtain, using lemma 3 and 7 in [13] and equations (12-15),

$$\|V_n - V^*\| \leq \|V^* - T^{in_0}(V_{n-in_0})\| + \|T^{in_0}(V_{n-in_0}) - V_n\|$$

$$\leq \|T^{in_0}(V^*) - T^{in_0}(V_{n-in_0})\| + \sum_{j=1}^{i} \|T^{jn_0}(V_{n-jn_0}) - T^{(j-1)n_0}(V_{n-(j-1)n_0})\|$$

$$\leq (\alpha_1)^i \|V^* - V_{n-in_0}\| + \sum_{j=1}^{i} (\alpha_1)^{j-1} \|T^{n_0}(V_{n-jn_0}) - V_{n-(j-1)n_0}\|$$

$$\leq (\alpha_1)^i \|\bar{V}-\underline{V}\| + \sum_{j=1}^{i} (\alpha_1)^{j-1} \sum_{r=1}^{n_0} \|T^r(V_{n-(j-1)n_0+r}) - T^{r-1}(V_{n-(j-1)n_0+r-1})\|$$

$$< \varepsilon/2 + \frac{1}{1-\alpha_1} n_0 \frac{(1-\alpha_1)\varepsilon}{2n_0} = \varepsilon$$

Since ε is arbitrary, this completes the proof.

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

$$\|V_1 - V\| \leq \|T(V)-V\| + \|T(V)-V_1\|$$

$$\leq \|T(V)-V\| + \max(\delta^-, \delta^+).$$

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

$$\varepsilon^- + \frac{\alpha_1}{1-\alpha_1}(\|(V-V_1)^+\| + \varepsilon^- + \|(V_1-V)^+\| + \varepsilon_0/\alpha_1).$$

This expression tends to 0, and will thus become smaller than ε after a finite number of iteration. ■

REFERENCES

[1] A. Alj and A. Haurie, "Hierarchical control of a population process with application to group preventive maintenance", Proceedings IFAC 2nd Symposium on Large Scale Systems Theory and Applications, Toulouse, France, pp. 24-26, June 1980.

[2] R. Bellman and S.E. Dreyfus, Applied Dynamic Programming, Princeton, New-Jersey, 1962.

[3] C. Berger, "Commande optimale de systèmes stochastiques markoviens de grande dimension", les cahiers du GERAD, no. G-80-04, Ecole des H.E.C., Montréal, 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, Academic Press, 1978.

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

[7] E.V. Denardo, "Contraction mappings in the theory underlying dynamic programming", SIAM Review, 9, pp. 165-177, 1967.

[8] B.L. Fox, "Discretizing dynamic programs", J. Optim. Theory Appl., II, pp. 228-234, 1973.

[9] I.B. Gertsbakh, Models of Preventive Maintenance, 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, Lect. Notes in Oper. Res. and Math. Syst., Springer Verlag, 1970.

[12] K. Hinderer, "On approximate solutions of finite stage dynamic programs", in Dynamic Programming and its Applications, M.L. Puterman ed., Academic Press, pp. 289-317, 1978.

[13] H.J. Langen, "Convergence of dynamic programming models", Math. of Oper. 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", les cahiers du GERAD, no. G-83-06, Ecole des H.E.C., Montréal, 1983.

[15] P. L'Ecuyer and A. Haurie, "Discrete event dynamic programming in Borel spaces with state dependent discounting", Dept. Informatique, U. Laval, Québec, Report no. DIUL-RR-8309, 1983.

[16] J. MacQueen, "A modified dynamic programming method for markovian decision problems", J. Math. Anal. Appl., 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., pp. 53-90, Academic Press, 1978.

[19] E. Porteus, "Some bounds for discounted sequential decision processes", Man. Sci., 18, pp. 7-11, 1971.

[20] E. Porteus, "Bounds and transformations for discounted finite markov decision chains", Oper. Res., 23, pp. 761-784, 1975.

[21] E. Porteus, "On the optimality of structured policies in countable stage decision processes", Man. Sci., 22, pp. 148-157, 1975.

[22] R. Rishel, "Group preventive maintenance: an example of controlled jump processes", Proceedings of the 20th IEEE Conference on Decision and Control, pp.786-791, San Diego, December 1981.

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

[24] W. Whitt, "Approximation of dynamic programs I and II", Math. of Oper. Res., vol. 3, pp. 231-243, 1978 and vol. 4, pp. 179-185, 1979.

[25] P. Whittle, "A simple condition for regularity in negative programming", J. Appl. Prob., 16, pp. 305-318, 1979.