

AD-A078 509

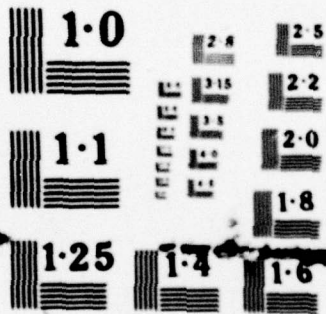
CALIFORNIA UNIV LOS ANGELES DEPT OF SYSTEM SCIENCE F/G 12/1
COUNTABLE STATE AVERAGE COST REGENERATIVE STOPPING PROBLEMS. (U)
APR 79 B L MILLER N00014-76-A-0004
UNCLASSIFIED UNCL A-WP-288 NL

UNCLASSIFIED

1 OF 1
AD-A078509



END
DATE
FILMED
1-80
DDC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

LEVEL 4

ADA078509

THE RUTH E. HOOVER
TECHNICAL LIBRARY
JUN 17 1979
NAVAL RESEARCH LABORATORY

JUN 22 1979
JUL 9 1979

COUNTABLE STATE AVERAGE COST REGENERATIVE
STOPPING PROBLEMS

by

BRUCE L. MILLER

JUL 19 1979

April 1979

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

DDC
RECEIVED
DEC. 19 1979
A

DDC FILE COPY

WESTERN MANAGEMENT SCIENCE INSTITUTE
University of California, Los Angeles

79 12 17 020

6 COUNTABLE STATE AVERAGE COST REGENERATIVE
STOPPING PROBLEMS

9 Working papers
by

10 Bruce L. Miller[†]

Department of System Science

University of California, Los Angeles

14
UCLA-WP-288

11 April 1979

12 31

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

DDC
RECEIVED
DEC 19 1979
A

[†] I thank Steve Lippman for reading an earlier draft and making a number of useful comments. I also thank Arie Hordijk for suggesting reference [3]. This research was supported in part by the Office of Naval Research under Contract N0014-76-A-0004 and the National Science Foundation under Contract ENG-76-12250.

15 N0014-76-A-0004,
-NSF-ENG 76-12250

405 213

mt

Abstract

Regenerative stopping problems are stopping problems which recommence from the initial state upon stopping. An algorithm is presented which solves a semi-Markov regenerative stopping problem with a finite number of continue actions by solving a sequence of stopping problems. New results for the optimal stopping problem are obtained as well as for the regenerative stopping problem. Two models in the literature are used as detailed examples of the algorithm.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DOC TAB	
Unannounced Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or special
A	

I. Introduction

Perhaps the most interesting result in the theory of optimal stopping from a computational standpoint is the monotone stopping theorem (Chow, Robbins, Siegmund [6, Theorem 3.3], Ross [14, Theorem 6.14], Abdel-Hameed [1]). That theorem, which gives conditions for which a myopic policy is optimal, is reconsidered in the next section. The main two differences from earlier versions are a more general cost structure and permitting a finite number of continue actions rather than one continue action. Our motivation for reconsidering the monotone stopping theorem is to apply it in solving regenerative stopping problems.

Regenerative stopping problems are stopping problems which recommence from the initial state upon stopping. They have an infinite planning horizon and either averaging or discounting must be used. The most important examples of regenerative stopping problems come from the literature on maintenance models, and a comprehensive description of maintenance models is available in the survey paper of Pierskalla and Voelker [16]. The impetus for this study came from Kaplan's model [14] of the optimal investigation of a production system. There the problem is to decide, based on reported monthly operating costs, when management should investigate and correct if necessary (stop). Once correction takes place the problem recommences from the initial state. In [5] Buckman and Miller solve the Kaplan model as a discounted regenerative stopping problem and also obtain some general results for discounted regenerative stopping problems.

Regenerative stopping problems were first studied independently by Brender [4] and Breiman [3]. Breiman called them binary decision renewal problems. Both authors proved that regenerative stopping problems could be solved by solving an appropriate stopping problem which we will call a λ -stopping problem

where λ has the interpretation as the average cost per period. The λ -stopping problem is defined by changing the cost associated with a continue action by subtracting the amount λ from it. Let $V(\lambda)$ be the expected cost of the λ -stopping problem using an optimal policy and starting from the initial state. The theorem of Brender and Breiman is that if λ^* satisfies $V(\lambda^*) = 0$, then the right λ has been used and the optimal decision rule for the λ^* -stopping problem is also the optimal decision rule for the regenerative stopping problem. For example, Taylor [21, Section 4] uses this theorem to solve an optimal replacement under cumulative damages problem by determining λ^* in closed form. In [11] Feldman has reconsidered a more general version of Taylor's problem and solved it by a different method. In problems where λ^* cannot be determined in closed form, an alternative would be to solve regenerative stopping problems by solving a sequence of λ -stopping problems ending with the λ^* -stopping problem, but neither Brender nor Breiman considered this approach. This approach seems quite promising if the monotone stopping theorem can be applied to each λ -stopping problem.

In Section 3 the theorem of Brender and Breiman is generalized by allowing a finite number of continue actions and letting the time spent in each state be a random variable so that the problem is semi-Markov. Three further results are obtained which lead to an algorithm for solving regenerative semi-Markov stopping problems by solving a sequence of λ -stopping problems.

In the last part of the paper the replacement-stockage model of Derman and Lieberman [9] and the maintenance model with uncertain information of Rosenfield [18] are solved by this algorithm, and the approach seems quite efficient and flexible. However the algorithm has not been tested by solving large scale problems.

II. Optimal Stopping

Our approach is to follow the formulation of Ross [20] and describe the optimal stopping problem as a Markov decision problem with a countable number of states $0, 1, \dots$ where state 0 is the initial state. Our formulation is discrete-time and the process is observed at time points $t = 0, 1, 2, \dots$. When the system is observed in state i at time t we choose from a finite set of continue actions A_i or decide to stop. If action $a \in A_i$ is chosen, we receive a cost of $C(i, a)$ and the process goes on to next state at time $t + 1$ according to the probabilities $P_{ij}(a)$. If we stop we receive the cost $C(i, s)$ and go to state Δ where we stay forever and $C(\Delta) = 0$ each period. The artificial state Δ is a notational convenience which allows us to let the planning horizon to be infinite. An admissible policy π is a decision rule which assigns to each state i and period t an action $\pi(i, t) \in A_i \cup \{s\}$, where s is the stop decision. Our objective is to find a decision rule which minimizes the expected cost up to and including stopping where the initial state is 0.

In order that our objective function be well-defined and that the monotone stopping theorem applies to our problem, we need some additional restrictions on the cost structure. We will assume that there is a scalar M and a set S containing Δ which satisfies the three assumptions below. Often S^c , the complement of S , will be a finite set of states and it may be empty. If S^c is not empty it will contain the initial state 0. Loosely speaking, the system starts in S^c and eventually reaches S , the "well-behaved" set (Assumption 3iii below).

Assumption 1. Either (i) or (ii) holds. Condition (i) is that $|C(i, a)| \leq M$ for all states i . Condition (ii) is that for all states i , $C(i, s) \geq -M$, and $P_{ij}(a) > 0$ implies that $C(j, s) \geq C(i, s)$. Furthermore if $i \in S^c$ then

$C(i,s) \leq M$, and $P_{ij}(a) > 0$ implies $C(j,s) \leq M$.

Assumption 2. $\inf_{i \in S, a \in A_1} (\min C(i,s)) \geq -M$, and $\sup_{i \in S^c, a \in A_1} (\max C(i,s)) \leq M$.

Assumption 3i. There are numbers N and $\delta > 0$ such that

$$\inf_{i \in S^c} \left\{ \sum_{j \in S} P_{ij}^N \right\} \geq \delta,$$

where P_{ij}^N is the probability of going from state i to state j in N periods and depends on the decision rule. We require that the above inequality holds for all decision rules.

(31i). The set S is closed. By this we mean that if $i \in S$ and $j \in S^c$ then $P_{ij}(a) = 0$ for all $a \in A_1$. The stop action also satisfies (ii) since $\Delta \notin S^c$.

(31ii). For some $\epsilon > 0$, $C(i,s) \geq \epsilon$ for all $i \in S \setminus \{\Delta\}$ and $a \in A_1$.

The Assumptions 1 and 2 imply that the costs are bounded above and below for states in S^c , and Assumption 3i says that there is an N -stage contraction on the probability of staying in S^c . Assumption 3ii assumes that when the system reaches the set S it will stay in S . Assumption 3iii states that there is a strictly positive cost of continuing when the system is in the set S .

Let Z_t and a_t be the state and action at time t . Then the expected return starting from state i and using the policy π is

$$G_\pi(i) = E_\pi \left(\lim_{n \rightarrow \infty} \left(\sum_{t=0}^n C(Z_t, a_t) \mid Z_0 = i \right) \right).$$

This expression is well-defined because of

Lemma 1. Let $x^- = \max(0, -x)$. Then for all policies π ,

$$E_\pi \left(\lim_{n \rightarrow \infty} \left(\sum_{t=0}^n C(Z_t, a_t)^- \mid Z_0 = i \right) \right) \leq (MN/\delta) + M.$$

Proof. The lemma follows from the fact that the continue cost per period is only negative when the system is in S^c , and the expected time in S^c is less than or equal to N/δ for any policy. To MN/δ we add M since $-M$ is a lower bound on the cost of stopping. Q.E.D.

We let $G(i) = \inf_{\pi} G_{\pi}(i)$ for all i , so that G is the optimal return function. We will prove (Lemmas 2 and 3) that G satisfies the equation of optimality

$$G(i) = \min \left(C(i,s), \min_{a \in A_1} (C(i,a) + \sum_j P_{1j}(a)G(j)) \right). \quad (1)$$

Our formulation differs from Ross [20] in two major ways. Unlike Ross we permit a finite number of continue actions. The generalization to a finite number of continue actions does not complicate the derivation of the monotone stopping theorem which gives a condition for stopping to be optimal. It does represent a major complication for obtaining the optimal policy since we must determine which continue action to use for states where stopping is not optimal. This complication is addressed in the last section where specific models are solved. It seems possible to further generalize the action space using, for example, the methods of Fox [12].

The other difference between our formulations and that of Ross is that he has more restrictions on the cost structure. He requires for all states i that (a) $0 \geq C(i,s) \geq -M$ and (b) for some $\epsilon > 0$, $C(i,a) \geq \epsilon$. Although on [20, p. 135] Ross does not require that $C(i,s) \leq 0$, his proof of the monotone stopping theorem requires nonpositive stopping costs.

The general statement of the optimal stopping problem is given in Chapter 3 of Chow, Robbins, and Siegmund [6]. They assume that a sequence of random variables Y_1, Y_2, \dots having a known joint distribution are observed. If we stop at

the n th stage after having observed y_1, \dots, y_n , then a cost of $x_n = f(y_1, \dots, y_n)$ is incurred. The objective is to stop so as to minimize the expected value of the cost received upon stopping. If for some sample path we never stop, then the cost is undefined so that this formulation requires that an admissible decision rule stop with probability one.

Beside the limitation of a countable state space, our Markov decision formulation is more restrictive than that of Chow, Robbins, and Siegmund [6]. For example, Derman and Sacks [10] consider an equipment replacement problem which fits our Markov decision formulation except that their criterion is to minimize the expected cost up to and including stopping divided by the number of periods until stopping. This cost structure can be handled by the Chow, Robbins, and Siegmund formulation but not by ours. In that paper they also mention the more plausible criterion of the expected cost up to and including stopping divided by the expected number of periods before stopping which is a regenerative stopping problem.

Returning to our model, we want to establish the monotone stopping theorem under Assumptions 1-3. Our approach follows that of Ross [20]. Although we admit policies where the expected time until stopping is infinite, we begin by observing that we can eliminate those policies from further consideration since their expected cost is infinite. This is true using Lemma 1, the fact that the expected time spent in $S \setminus \{\Delta\}$ is infinite, and Assumption 3iii. In the lemmas to follow we will implicitly use the fact that $G(\Delta) = 0$.

Lemma 2. For $j \in S \setminus \{\Delta\}$,

- (a) The optimal return function satisfies the equation of optimality (1).
- (b) The stationary policy which for each state $j \in S$ selects the action which minimizes the right hand side of (1) is optimal.

(c) The optimal return function G satisfies $C(j,s) \geq G(j) \geq -M$.

Proof. The set S is closed and the states in S satisfy Ross' restrictions on the cost structure. Furthermore, as noted above we only need consider decision rules such that the expected time until stopping is finite. Therefore Ross' arguments [20, p. 135-136] apply directly to prove (a) and (b) of the lemma. The upper bound in (c) is obvious and the lower bound holds since the cost of stopping is bounded below by $-M$ and the costs of continuing are nonnegative. Q.E.D.

Lemma 3. For $j \in S^c$,

(a) and (b) of Lemma 2 hold.

(c) The optimal return function G is bounded above and below.

Proof. Let u and v be bounded functions on S^c , π be a policy defined on S^c and $\rho(u,v) = \sup_{i \in S^c} |u(i) - v(i)|$. Let $T_\pi^N(u)$ be the expected return in the first N periods using the policy π where u is the terminal reward vector, under the assumption that if a state $j \in S$ is reached in some period n , $n \leq N$, then the process terminates with a reward $G(j)$. Therefore $\rho(T_\pi^N(u), T_\pi^N(v)) \leq (1-\delta) \rho(u,v)$ since expected probability of leaving S^c in N periods is at least δ by Assumption 3i. Also $T_\pi^N(\cdot)$ is bounded if u is bounded since both the costs of continuing and stopping are bounded on S^c , and the possibility of reaching S does not destroy the boundedness properly. The latter is established by combining Lemma 2c with the implication of Assumption 1ii that the first transition out of S^c will be to a state with a bounded stopping cost. Thus we can have the N -state contraction property, and (a), (b), and (c) of the lemma follow from Denardo [7]. Q.E.D.

Lemma 4. The nonnegative term $C(i,s) - G(i)$ is bounded.

Proof. If Assumption 1i holds the result is immediate since $C(i,s)$ is bounded above, and by Lemmas 2c and 3c $G(i)$ is bounded below.

If Assumption 1ii holds then for $i \in S$ $C(i,s) = G(i)$ since the continue costs are nonnegative and $C(j,s) \geq C(i,s)$ if $P_{ij}^a \geq 0$. For $i \in S^c$ $C(i,s)$ is bounded above by assumption, and by Lemma 3c $G(i)$ is bounded below. Q.E.D.

Let G^n be the optimal return function for an n period problem where we are required to have stopped after n periods. Clearly $G^n \geq G^{n+1} \geq G$.

Lemma 5. For each state i , $\lim_{n \rightarrow \infty} G^n(i) = G(i)$.

Proof. By Lemmas 2b and 3b there is an optimal policy which can be obtained from (1). Following Ross [20, Theorem 6.13] we let π be this policy and π_n be the policy which uses the same action as π for periods $0, 1, \dots, n-1$, but stops in period n . Then $G^n(i) \leq G_{\pi_n}^n(i)$ and

$$G_{\pi_n}^n(i) - G(i) = \sum_{j=0}^{\infty} [C(j,s) - G(j)] P[Z_n=j]$$

where Z_n is the state in period n . By Lemma 4 the term in brackets is bounded and since the expected time until stopping starting from any state i is bounded, $P[Z_n=\Delta] \rightarrow 1$ as $n \rightarrow \infty$ which show that $\lim_{n \rightarrow \infty} G_{\pi_n}^n(i) = G(i)$. Q.E.D.

Lemma 5 has been called a stability condition by Ross [20] and Breiman [3].

Lemmas 2, 3, and 5 are sufficient to establish the monotone stopping theorem.

Let

$$B = \left\{ i: C(i,s) \leq \min_{a \in A_i} (C(i,a) + \sum_j P_{ij}^a C(j,s)) \right\}.$$

B is precisely those states for which stopping is at least as good as continuing one more period and then stopping. Clearly if $i \notin B$ we continue, but not necessarily with the action which minimizes $C(i,a) + P_{ij}^a \sum C(j,s)$. Rather than stating the monotone stopping theorem for the set B we state it for a subset of B . This can be useful since a subset of B may be easier to identify than B .

The Monotone Condition. A set of states D satisfies the monotone condition if it is closed and is a subset of B .

Theorem 1. (The Monotone Stopping Theorem). If a set D satisfies the monotone condition then for $i \in D$ the optimal decision is to stop.

Proof. The proof of Ross [20, Theorem 6.14] applies. Using the definition of B and the fact that $D \subset B$ is closed, Ross' proof shows that $G^n(i) = C(i, s)$ for all n and $i \in D$. Then by Lemma 5 $G(i) = \lim_{n \rightarrow \infty} G^n(i) = C(i, s)$. Therefore the stop decision minimizes the right hand of side of eq. (1), and by Lemmas 2b and 3b the stop decision is therefore optimal. Q.E.D.

III. Regenerative Stopping Problems

Regenerative stopping problems are stopping problems which return to state 0 upon stopping and recommence. We will continue with the countable state space of the previous section except for eliminating the artificial state Δ . However the transition times will be generalized so that we have a semi-Markov formulation. There is no standard semi-Markov decision model (compare for example Denardo [8], Lippman [15], and Ross [19]) and we will use one of the simpler versions.

The process begins at state 0 at time 0. When the system is observed in state i (immediately) after the n th transition, we choose from a finite set A_i of continue actions or decide to stop. If the action $a \in A_i$ is chosen the process goes to the next state according to the probabilities $P_{ij}(a)$. Given that the next state j , the time for the transition to take place is a random variable T_{ija} . We will require that for some $\delta > 0$, $\epsilon > 0$, $\sum_{j=0}^{\infty} P_{ij}(a) F_{ija}(\delta) \leq 1 - \epsilon$ for all i and a where $F_{ija}(\cdot)$ is the distribution function of T_{ija} . This condition says that for every state i and continue action a there is a positive probability of at least ϵ that the transition time will be greater

than δ which, in turn, means that there is a strictly positive bound on the expected time until the system returns to state 0. For any state i , $i \geq 0$, we may choose to stop. Then $P_{i0}(s) = 1$ and the time until reaching state zero is given by the random variable T_{is} . There are no restrictions on T_{is} except when $i = 0$. For state 0 we require that for some $\delta > 0$, $\epsilon > 0$, $\text{Prob}(T_{0s} > \delta) > \epsilon$. We require that both $E[T_{ija}]$ and $E[T_{is}]$ be finite. The costs for continuing and stopping are given by the nonnegative random functions $C(i,s,T_{ija})$ and $C(i,s,T_{is})$ respectively. They are both incurred at the beginning of a transition. Following Lippman [15] we will let a policy π be a decision rule which, given the number of the transition and the state, says which action is to be chosen. It would be preferable to have the decision rule depend on the time of the transition rather than the number of the transition, since in a finite horizon semi-Markov problem the optimal policy would depend on the time remaining to the end of the horizon but not the number of the transition (Jewell, [13]). However, the additional complication of allowing the decision rule to depend on time does not seem justified for our purposes. For a policy π we let

$$X_{\pi} = \limsup_{t \rightarrow \infty} \frac{1}{t} E[W_{\pi}(t)]$$

where $W_{\pi}(t)$ is the expected cost up to and including t using the policy π starting from state 0, so that X_{π} represents the average cost per period using the policy π . The conditions on the T_{ija} and T_{0s} and the nonnegativity of the costs assure that X_{π} is well-defined although possibly infinite. Let $X^* = \inf_{\pi} X_{\pi}$. Our objective is to find a policy π^* such that $X_{\pi^*} = X^*$.

We propose to solve our semi-Markov regenerative stopping by solving a sequence of λ -stopping problems where λ , $-\infty < \lambda < \infty$, has the interpretation as the average cost per period. A λ -stopping problem is constructed from the

data of a semi-Markov regenerative stopping problem by letting the $P_{1j}(a)$ and the time between transitions remain unchanged, and setting

$$C(i,a) = \sum_j P_{1j}(a) E[C(i,a,T_{1ja}) - \lambda T_{1ja}]$$

and

$$C(i,s) = E[C(i,s,T_{1s}) - \lambda T_{1s}].$$

Thus we have an optimal stopping problem which may or may not satisfy Assumptions 1-3, and for which the length of time between transitions is a random variable. Clearly the optimal stopping problem is unaffected by the length of time between transitions as long as they are finite with probability one.

Let $G(i,\lambda)$ be the optimal return function of the λ -stopping problem from the initial state i . The case where $i = 0$ is important enough that we introduce the function V defined by $V(\lambda) = G(0,\lambda)$. Let Λ be the set of λ such that the λ -stopping problem satisfies Assumptions 1-3 of the previous section. The set Λ is a semi-infinite interval since if $\lambda' \in \Lambda$ and $\lambda \leq \lambda'$ then $\lambda \in \Lambda$ since only Assumption 3iii will depend on the choice of λ and it is easier to satisfy the smaller the value of λ . Unless otherwise stated, we will be assuming that $\lambda \in \Lambda$ for any λ -stopping problem being considered. Besides illustrating the notation, the following is an example of a problem where there is no λ^* such that $V(\lambda^*) = 0$.

Example 1. Consider a discrete time problem where there is only one continue action for each state, and $P_{1,1+1}(a) = 1$ and $C(i,a,1) = 1$ for all i . For states $i \geq 1$, the stop action takes no time ($T_{1s} = 0$ with probability one) and $C(i,s,0) = 1$. For state 0 the stop action takes one period at a cost of 10. By inspection the optimal policy for the regenerative stopping problem is to never stop and the average cost per period is 1. If $\lambda < 1$ then the set S can

be the entire space $0,1,2,\dots$ and Assumptions 1-3 holds for the λ -stopping problem. If $\lambda \geq 1$, the set S must be empty by Assumption 3iii, but then Assumption 3i cannot be satisfied. Therefore $\Lambda = (-\infty, 1)$.

The function $V(\lambda)$ for this problem is

$$V(\lambda) = 2 - \lambda \quad \lambda \leq 1 \quad (\text{the optimal policy is to stop after the transition from state 0 to state 1}).$$

$$V(\lambda) = -\infty \quad \lambda > 1 \quad (\text{the optimal policy is to never stop}).$$

Clearly there is no λ^* such that $V(\lambda^*) = 0$.

Let π be a policy for the λ -stopping problem which satisfies $E[T] < \infty$ where T is the random time until stopping using the policy π . Let $V_\pi(\lambda)$ be the expected cost of the λ -stopping problem starting from state 0 using the policy π . Then

$$V_\pi(\lambda) = E[C] - \lambda E[T] \quad (2)$$

where C is the original (without subtracting the λ terms) cost up to and including stopping using the policy π . Thus $E[C] = V_\pi(0) = G_\pi(0,0)$. Equation (2) simply decomposes the costs of a λ -stopping problem.

We will now prove the theorem of Brender [4] and Breiman [3] in the semi-Markov case where policies are admitted where the expected time until stopping is infinite.

Theorem 2. Suppose that $\lambda^* \in \Lambda$ satisfies $V(\lambda^*) = 0$. Then the policy which is optimal for the λ^* -stopping problem is optimal for the regenerative stopping problem. Also $\lambda^* = X^*$, the optimal expected cost per period.

Proof. Let π be a stationary policy which solves the λ^* -stopping problem. Such a policy exists by Lemmas 2 and 3 and furthermore $E[T_\pi] < \infty$ where T_π is the random time until stopping using the policy π . The expected cost up to and including returning to state 0 using the policy π is $\lambda^* E[T_\pi]$ from (2) since $V_\pi(\lambda^*) = V(\lambda^*) = 0$.

Let Y_1 be the random cost in the regenerative stopping problem for the i th return to state 0 using the policy π . We have just shown that $E[Y_1] = \lambda^* E[T_\pi]$. By Ross [20, Theorem 3.16] including his subsequent remarks $\lim_{t \rightarrow \infty} \frac{1}{t} E[W_\pi(t)] = \lambda^*$. Since $X_\pi = \lim_{t \rightarrow \infty} \frac{1}{t} E[W_\pi(t)]$ we will have established that $\lambda^* = X_\pi$ once we have established that π is the optimal policy for the regenerative stopping problem.

Now let π' be an arbitrary admissible policy for the regenerative stopping problem with $E[T_{\pi'}] < \infty$. The expected cost up to and including returning to state 0 using the policy π' for the regenerative stopping problem is greater than or equal to $\lambda^* E[T_{\pi'}]$ using $V_{\pi'}(\lambda^*) \geq V(\lambda^*) = 0$ and equation (2). The same analysis as above shows that $\lim_{t \rightarrow \infty} \frac{1}{t} E[W_{\pi'}(t)] \geq \lambda^*$.

Let π' be an arbitrary policy for the regenerative stopping problem with $E[T_{\pi'}] = \infty$. We consider a modified regenerative stopping problem where

$$C(i, a, T_{ija})' = C(i, a, T_{ija}) - \lambda^* T_{ija}$$

and

$$C(i, s, T_{is})' = C(i, s, T_{is}) - \lambda^* T_{is}$$

The proof consists of showing that the average reward of the modified regenerative stopping problem is greater than or equal to zero and therefore that the average reward for the original regenerative stopping problem is greater than or equal to λ^* .

The expected cost up to and including returning to state 0 for the modified regenerative problem is $V_{\pi'}(\lambda^*)$ since the modified regenerative problem has the same costs as the λ^* -stopping problem. By using Lemma 1 ($\lambda^* \in \Lambda$) and the fact that $E[T_{\pi'}] = \infty$, Ross [20, Theorem 3.16] can be easily modified to show that the average reward of the modified regenerative stopping problem is greater than or equal to zero. Q.E.D.

The following propositions will be used in the algorithm for solving regenerative stopping problems. The first establishes some useful properties

of the V function. The second gives an alternative optimality condition, and the third shows how the solutions improve as successive λ -stopping problems are solved.

Proposition 1. $V: \Lambda \rightarrow \mathbb{R}$ is a decreasing, and finite-valued, and concave function of λ . Since V is concave it is known that the right and left hand derivatives exist everywhere on the interior of Λ . Furthermore

$$V'_-(\lambda) \geq -E[T_\lambda] \geq V'_+(\lambda) \quad (3)$$

where V'_- and V'_+ are the left and right hand derivatives of V and T_λ is the stopping time of a λ -optimal policy.

Proof. It is clear that V is decreasing in λ . Furthermore $V > -\infty$ for $\lambda \in \Lambda$ since Lemma 1 applies. To show that V is concave consider points $\lambda, \alpha\lambda_1 + (1-\alpha)\lambda_2$, and λ_2 where $0 < \alpha < 1$. Let π be an optimal stationary policy for the $\alpha\lambda_1 + (1-\alpha)\lambda_2$ -stopping problem. If that same policy π is used for the λ_1 and λ_2 stopping problem then from (2)

$$V(\lambda_1\alpha + (1-\alpha)\lambda_2) = V_\pi(\alpha\lambda_1 + (1-\alpha)\lambda_2) = \alpha V_\pi(\lambda_1) + (1-\alpha) V_\pi(\lambda_2).$$

However

$$V_\pi(\lambda_1) \geq V(\lambda_1) \text{ and } V_\pi(\lambda_2) \geq V(\lambda_2) \text{ which shows the concavity of } V.$$

The inequalities on the right and left hand derivatives are established by using a similar approach. Let π be the optimal policy for the λ -stopping problem. Then for $\epsilon > 0$,

$$V(\lambda+\epsilon) - V(\lambda) \leq V_\pi(\lambda+\epsilon) - V_\pi(\lambda) = -\epsilon E[T_\pi]$$

where T is stopping time of the policy π . Letting ϵ go to zero establishes the result for $V'_+(\lambda)$. The proof for $V'_-(\lambda)$ is similar. Q.E.D.

Proposition 2. If π is optimal for both a λ_1 -stopping problem and a λ_2 -stopping problem, where $V(\lambda_1) \geq 0$ and $V(\lambda_2) \leq 0$, $\lambda_1 < \lambda_2$, and $\lambda_1, \lambda_2 \in \Lambda$, then π is optimal for the regenerative stopping problem.

Proof. From (2) $V_{\pi}(\lambda) = V_{\pi}(\lambda_2) - (\lambda - \lambda_2) E(T_{\pi})$. By Proposition 1, $V(\cdot)$ is concave, $V'_{-}(\lambda_2) \geq -E(T_{\pi})$, and $V'_{+}(\lambda_1) \leq -E(T_{\pi})$. Therefore $V'(\lambda) = -E(T_{\pi}) = V'_{\pi}(\lambda)$ for $\lambda_1 < \lambda < \lambda_2$. This together with either $V_{\pi}(\lambda_1) = V(\lambda_1)$ or $V_{\pi}(\lambda_2) = V(\lambda_2)$ implies that $V_{\pi}(\lambda) = V(\lambda)$ for $\lambda_1 \leq \lambda \leq \lambda_2$. Furthermore since $V(\lambda_1) \geq 0$ and $V(\lambda_2) \leq 0$, it is clear that $V(\lambda) = 0$ for some λ between λ_1 and λ_2 and that π is optimal for the regenerative stopping problem by Theorem 2. Q.E.D.

Proposition 3. If $V(\lambda_2) > V(\lambda_1) \geq 0 = V(\lambda^*)$ then the λ_1 -optimal policy is at least as good as the λ_2 -optimal policy with respect to the regenerative stopping problem. Likewise if $V(\lambda_2) < V(\lambda_1) < 0 = V(\lambda^*)$ the λ_1 -optimal policy is at least as good as the λ_2 -optimal policy with respect to the regenerative stopping problem.

Proof. We only prove the proposition for the case $V(\lambda_2) > V(\lambda_1) > 0$ since the other proof is similar. Let π_1 be the λ_1 -optimal policy and π_2 be the λ_2 -optimal policy.

First we establish the inequality $E[T_2] \leq E[T_1]$ where the subscript 1 refers to the policy π_1 and 2 refers to π_2 . Assume the contrary. Then $(\lambda_2 - \lambda_1) E[T_2] < (\lambda_2 - \lambda_1) E[T_1]$ since $\lambda_2 < \lambda_1$. Also $E[C_2] - \lambda_2 E[T_2] \leq E[C_1] - \lambda_2 E[T_1]$. Adding these two inequalities implies that π_2 is strictly better than π_1 for the λ_1 -stopping problem, a contradiction.

Returning to the main argument,

$$E[C_2] - \lambda_1 E[T_2] \geq E[C_1] - \lambda_1 E[T_1] = V(\lambda_1) > 0$$

If we divide both sides by $E[T_2]$ and $E[T_1]$ respectively, the inequality is maintained and

$$\frac{E[C_2]}{E[T_2]} - \lambda_1 \geq \frac{E[C_1]}{E[T_1]} - \lambda_1. \quad \text{Q.E.D.}$$

From example 1 we see that for some problems there is no λ such that $V(\lambda) = 0$. In that example the optimal policy was to never stop. Proposition 1

implies that the existence of a $\bar{\lambda} \in \Lambda$ such that $V(\bar{\lambda}) < 0$ is a sufficient condition that there is a $\lambda^* \in \Lambda$ such that $V(\lambda^*) = 0$. The existence of such a $\bar{\lambda}$ is often easy to verify. If we have such a $\bar{\lambda}$ then the algorithm below will only consider $\lambda \leq \bar{\lambda}$ and which therefore belong to Λ .

The Regenerative Stopping Algorithm.

Step O.A. Find a λ_0 which is less than X^* , the optimal average cost of the regenerative stopping problems. The models of the next section will provide examples of how this λ_0 can be found. It is desirable that λ_0 be as large as possible. We solve the λ_0 -stopping problem and let π be the optimal policy for that problem. Since $\lambda_0 \leq X^*$, $V(\lambda_0) \geq 0$. If a mistake is made and λ_0 is greater than X^* , then $V(\lambda_0) < 0$ and λ_0 can be changed until $V(\lambda_0) \geq 0$.

Step O.B. Set $\lambda_1 = \min(X_\pi, \bar{\lambda})$ where π is the optimal policy of the λ_0 -stopping problem. We solve the λ_1 -stopping problem. Since $\lambda_1 \geq X^*$, $V(\lambda_1) \leq 0$. We check if Theorem 2 or Proposition 2 is satisfied. If not we continue to Step 1.

Step 1. We are now in the general case where we have solved a λ_0 -stopping problem and a λ_1 -stopping problem where $\lambda_0 \leq X^*$ and $\lambda_1 \geq X^*$. The new λ -stopping problem to be solved is given by $\lambda^{\text{new}} = \min(\lambda, X_\pi)$ where π is the best (lowest average cost) policy determined to date, and $\lambda = \alpha\lambda_B + (1-\alpha)\lambda_A$ where $0 < \alpha \leq 1$. The subscript B stands for bisection and the subscript A stands for approximation. Computational experiences suggests choosing a low value of α , since the approximation is quite accurate. We have $\lambda_B = 1/2 \lambda_0 + 1/2 \lambda_1$. λ_A is the λ such that $V_A(\lambda) = 0$, where $V_A(\lambda)$ is based on the four equations:

$$\begin{aligned} V_A(\lambda_0) &= V(\lambda_0) & , & & V'_A(\lambda_0) &= -E[T_{\lambda_0}] , \\ V_A(\lambda_1) &= V(\lambda_1) & , & & \text{and } V'_A(\lambda_1) &= -E[T_{\lambda_1}] . \end{aligned} \tag{4}$$

These equations determine the coefficients of the cubic approximation $V_A(\lambda) = B_0 + B_1\lambda + B_2\lambda^2 + B_3\lambda^3$. The derivative conditions are based on Proposition 1. The λ^{new} -stopping problem is solved and λ^{new} replaces λ_1 if $V(\lambda^{\text{new}}) < 0$ and replaces λ_0 if $V(\lambda^{\text{new}}) > 0$. We check if Theorem 2 or Proposition 2 is satisfied. If not we return to Step 1.

Comments. A value of $\alpha > 0$ in Step 1 assures that the "interval of uncertainty" goes to zero. Besides assuring that the "interval of uncertainty" goes to zero there is a rationale for a positive α even if the cubic approximation is excellent. Suppose that $\alpha = 0$ and λ_A is close to either λ_0 or λ_1 say λ_1 . Then we would prefer that $V(\lambda_A) \geq 0$ so that the next interval of uncertainty is (λ_A, λ_1) rather than (λ_0, λ_A) . If $\alpha > 0$ then $\lambda^{\text{new}} < \lambda_A$ and $V(\lambda^{\text{new}}) \geq V(\lambda_A)$ so that $V(\lambda^{\text{new}})$ is "more likely" than $V(\lambda_A)$ to be greater than or equal to zero. When λ_A is roughly between λ_0 and λ_1 the choice of α is not important.

If we stop before optimality Proposition 3 says that either the current λ_0 -optimal policy or the current λ_1 -optimal policy will be the best policy determined to date, and that earlier efforts can be forgotten. The higher of the average returns of these two policies can be compared with the current lower bound on X^* , λ_0 .

An alternative to Step 1 would be the policy iteration approach where $\lambda^{\text{new}} = X_\pi$, where π is the most recently considered policy. In this case the sequence of λ would be decreasing to X^* .

Finally, let π be the optimal policy of a λ -stopping problem. It is interesting to observe, using the notation of (2), that

$$X_\pi = \frac{E[C_\pi]}{E[T_\pi]} = \lambda + \frac{E[C_\pi] - \lambda E[T_\pi]}{E[T_\pi]} = \lambda + \frac{V(\lambda)}{E[T_\pi]} = \lambda - \frac{V(\lambda)}{V'(\lambda)}$$

when $V'(\lambda)$ exists, the last equality by Proposition 1. Therefore X_π equals the point where a supporting hyperplane at $V(\lambda)$ to the concave function V

would equal 0. This is reminiscent of Puterman and Branelle's result [17] relating Newton's method and policy iteration in the finite state finite action Howard model.

IV. A Replacement-Stockage Model

Derman and Lieberman [9] consider a machine which requires one transistor. When a new transistor is installed there is a probability f_s that it will perform at service level s , $s = 1, 2, 3, \dots$. After each period the service level either stays at the same level or the transistor fails with probability p , independent of the length of service. When the transistor is in service, at the end of the period one may either leave it in service or remove it if the service level is unsatisfactory. Spare transistors are kept in a bin according to the rule that, when empty, the bin is restocked with N new transistors and the machine is shut down for one period. The objective is to minimize expected average cost per period over an infinite horizon. The problem is to determine a restocking level N^* and a rule for replacing a transistor in service which meet this objective. For simplicity we will assume a less general cost structure than they did, but one which does include their example problem. We assume an ordering cost of $K + cN$ when $N \geq 1$ transistors are ordered where the constant K includes the cost of operating with zero transistors during the one period it takes for the order to be received. The operating cost per period is $hn + w_s$ where s is the level of service and n is the number of transistors available, $0 \leq n \leq N$, and h is a positive holding cost. We also want w_s nonnegative and increasing in s .

Derman and Lieberman [9] formulate the problem as a countable state Markov decision problem with states (i, s) $i \geq 1$, $s \geq 1$, and a state 0. When the system is in state (i, s) they mean that i units of stock are on hand of which one is installed at operating level s . The possible decisions in state (i, s) are to replace the unit in service at the end of the period or not to replace the

unit in service. At state 0 they decide how many units to order. On page 615 of [9] they outline their algorithm whose overall plan is to decompose the replacing and the ordering decisions. First they determine an upper bound \bar{N} for N , and for $N = 1, \dots, \bar{N}$, calculate by policy iteration the optimal policy for the problem of determining for which service levels the transistors should be replaced with a new one. Derman and Lieberman develop several tests to speed up the calculations. Bell [2] reconsiders the problem and applies the monotone stopping theorem to obtain some new tests to speed up the calculations, but the overall approach is that of Derman and Lieberman.

We will reformulate the problem by simplifying the state space at the price of enlarging the action space. We let the state space be the integers $i, i \geq 1$, where state i means that i units are on hand including any in service. In this formulation 1 is the initial state. For any state i there are a countable number of continue actions $a = 1, 2, 3, \dots, \infty$ where taking action a means that the installed unit will be replaced if it is operating at level a or worse. The expected cost for state i and action a is

$$C(i, a) = \sum_{s < a} (hi + w_s) f_s \frac{1}{p} + \sum_{s \geq a} (hi + w_s) f_s. \quad (5)$$

The expected length of time in state i is $\sum_{s < a} f_s \frac{1}{p} + \sum_{s \geq a} f_s$. For convenience we have assumed a highly plausible form of the replacement rule. This assumption can be justified from equation (6) which follows. Because of the simple way in which the costs and expected transition time vary with the action a , the countable number of decisions does not cause a computational difficulty. Since the best action can always be determined there is no theoretical complication from going from a finite number of continue actions to an infinite number.

We make a second major alteration in the formulation by assuming that $P_{ij}^a = 1$ if $j = i + 1$ rather than for $j = i - 1$. Thus if N is the reorder level we will perceive the stock level as going from the states 1 to 2, ... to N to $N + 1$

to 1, rather than as they do physically from N to $N-1, \dots$ to 0 , to N . In the averaging case this reordering will lead to the same average cost per period. The purpose in this reordering is to allow a stop decision with each state i , $i \geq 2$, and so that for the λ -stopping problems the set B will satisfy the monotone condition and Theorem 1 can be used. The cost of stopping, $C(i,s)$, is $K + c(i-1)$, it takes one period, and the system returns to state 1. We multiply c by $i - 1$ since if we stop at state i the reorder level is $i - 1$. For each λ -stopping problem the smallest state where we stop, and hence the reorder level, will be determined by applying the monotone stopping theorem.

To see that this reformulation is justified consider the policy which in the Derman and Lieberman formulation orders two units in state 0, replaces in state $(1,s)$ if $s \geq 3$, and replaces in state $(2,s)$ if $s \geq 2$. The expected cost per cycle is $K + 2c + \sum_{s \leq 2} (h+w_s) f_s \frac{1}{p} + \sum_{s \geq 3} (h+w_s) f_s + (2h+w_1) f_1 \frac{1}{p} + \sum_{s \geq 2} (2h+w_s) f_s$ and the expected length of the cycle is $1 + \sum_{s \leq 2} f_s \frac{1}{p} + \sum_{s \geq 3} f_s + f_1 \frac{1}{p} + \sum_{s \geq 2} f_s$. In our formulation that policy is: choose action 3 when in state 1, action 2 when in state 2, and stop when in state 3. It can easily be seen to have the same expected cost and same expected length per cycle as those just given.

In order to apply the Regenerative Stopping Algorithm, we first check Assumptions 1-3 for the λ -stopping problem in order to determine Λ . For any λ let $S = \{i: h_i > \lambda\}$. Assumptions 3i, 3ii, 3iii are satisfied. Assumption 2 is satisfied. Assumption 1i is not satisfied but Assumption 1ii is satisfied. Thus $\Lambda = (-\infty, +\infty)$ and clearly there is a large $\bar{\lambda} \in \Lambda$ such that $V(\bar{\lambda}) < 0$. Next we look at the set $B = \{i: C(i,s) \leq (\min_{s \in A_1} C(i,s) + C(i+1,s)), i \geq 2\} = \{i: 0 \leq \min_{s \in A_1} C(i,s) + c, i \geq 2\}$ where $C(i,s)$ is the cost of stopping for the

λ -stopping problem and equals $K + c(i-1) - \lambda$. The $C(i,a)$ are the continue costs for the λ -stopping problem and equal $\sum_{s < a} (hi + w_s - \lambda) f_s \frac{1}{p} + \sum_{s \geq a} (hi + w_s - \lambda) f_s$.

Since $C(i,a)$ is increasing in i because of the holding costs, B is of the form $\{i: i \geq i^*\}$ for some integer i^* which is a closed set and Theorem 1 can be applied. It turns out that the best continue decision is easily determined for the λ -stopping problem since $P_{i+1}(a) = 1$ regardless of the choice of a . For any state i the best continue action, $a^*(i)$, is

$$a^*(i) = \inf \{s: hi + w_s > \lambda\}. \quad (6)$$

It is precisely (6) that Bell [2] exploits in his approach to this problem. We are ready to apply the algorithm once we have determined λ_0 , the lower bound on X^* . This lower bound is obtained by assuming that $f_1 = 1$, that we always go to the most favorable operating state. This assumption eliminates the replacement decision and

$$\lambda_0 = \min_{N \geq 1} \left\{ \left(K + \sum_{i=1}^N [c + (w_1 + hi) \frac{1}{p}] \right) / (1 + N/p) \right\}. \quad (7)$$

Example 2. The first problem we consider is one that is presented in both Derman and Lieberman [9] and Bell [2]. The data are $p = .1$, $f_s = (1/2)^s$; $s = 1, 2, 3, \dots$. $h = 4$, $w_s = 100 (1.4 - (.2/2^{s-1})) - 4$, $K = 140$, and $c = 20$. In this example we will let α , the weighting factor of the Regenerative Stopping Algorithm be .1.

First eq. (7) is solved and $\lambda_0 = 123.63$ with the minimizing $N = 1$. Then the λ_0 -stopping problem is solved. For state 1 $a^*(1) = 2$ using (6) since $4 + 116 \leq 123.63$ but $4 + 126 > 123.63$. The value of $C(1,2) = (4 + 116 - 123.63) (1/2)(10) + \sum_{s \geq 2} (4 + w_s - 123.63) f_s = -13.29$. For state 2 $a^*(2) = 1$ and $C(2,1) \geq 0$ so that $2 \in B$. Since we stop at state 2 the reorder level is 1. $G(2, 123.63) = 160 - 123.63 = 36.37$, and $G(1, 123.63) = -13.29 + 36.37 = 23.08 = V(123.63)$.

We do not determine a value for $\bar{\lambda}$, and $\lambda_1 = 123.63 + (23.08/6.5) = 127.18$, the average cost per period of the above policy π . When we solve the 127.18-stopping problem, the policy π is again optimal and $V(127.18) = 0$ so that π is optimal by Theorem 2.

In order to gain more experience with the algorithm the following problems were solved: As before $p = .1$, $f_s = (1/2)^s$, $s = 1, 2, 3, \dots$, $K = 140$, $c = 20$, but $h = .1$ and $w_s = H(1 - (15/16)^{s-1}) - .1$ where H is a scalar. These problems had optimal order quantities N^* of between 10 and 20. Recall that the algorithms of Bell [2] and Derman and Lieberman [9] must solve N^* policy iteration problems.

Different runs were made by using different values of the weighting factor α and the parameter H . The computational results in terms of the number of λ -stopping problems solved were

	$\alpha = 0.0$	$\alpha = .02$	$\alpha = .10$	$\alpha = .25$
H = 4.6	4	5	6	6
H = 4.8	5	5	5	6
H = 5.0	5	5	5	6
H = 5.2	5	5	5	5

V. Markov Deterioration with Uncertain Information

Rosenfield [18] considers a maintenance problem where the underlying Markov process has actual states $0, 1, \dots, N$, where 0 is the best state and N is the worst state. The actual state is not known except at certain times and the observed state is (i, k) which means that k periods ago the system was observed in state i , where $0 \leq i \leq N$, and $k \geq 0$. Each period the actual state of the system changes according to a Markov transition matrix P . Thus if the process is in state (i, k) the probability the actual state is j is P_{ij}^k , the ij element of the matrix P to the k th power.

Each period there are three available decisions for state (i,k) , repair, no action, or inspect. If the repair decision is selected, the cost is R and the system moves to state $(0,0)$ the next period. With the decision no action the expected cost is $\sum_{j=0}^N P_{ij}^k L_j$ and the system moves to state $(i,k+1)$ the next period, where L_j is interpreted as the one-period operating cost when the machine is in state j . The expected cost associated with the inspect decision is $M + \sum_{j=0}^N P_{ij}^k L_j$ where M is the cost of inspection, and the system moves to state $(j,0)$ with probability P_{ij}^{k+1} .

This is the Rosenfield model except that we have not permitted the cost of repair to depend on the actual state. Like Rosenfield we will require that the matrix P satisfy $P_{ij} = 0$ if $j < i$, $P_{ij} < 1$ except for $j = N$, and $\sum_{j=k}^N P_{ij}$ is non-decreasing in i for $k = 1, 2, \dots, N$. Rosenfield cites results which show that if P satisfies the above conditions then so will P^k . We will assume that the L_j are increasing in j . From Rosenfield [18, Lemma 1] we have the results that for any increasing function W_j , $0 \leq j \leq N$, $\sum_{j=0}^N P_{ij}^k W_j$ is increasing in i and in k when P satisfies the above conditions.

As with the replacement and stockage model of the previous section we will reformulate the problem and simplify the state space by enlarging the action space. In our formulation the states will be the actual states $0, 1, \dots, N$. For each state i there will be both a countable number of continue actions and a countable number of stopping actions. The continue actions are of the form a -inspect, $a \leq 0$, and have the interpretation, "inspect after a periods have passed." The decision takes $a + 1$ periods and the associated cost when in state i is $(L_i + \sum_j P_{ij} L_j + \dots + \sum_j P_{ij}^a L_j + M)$. For $a = 0$ this cost is $L_i + M$. The system then moves to state j with probability P_{ij}^{a+1} . The stopping actions are of the form a -repair, $a \geq 0$, and have the interpretation, "repair after a periods have passed." This decision takes $a + 1$ periods and the associated cost

when in state i is $(L_1 + \sum_j P_{1j} L_j + \dots + \sum_j P_{1j}^{a-1} L_j + R)$. For $a = 0$ this cost is R , and for $a = 1$ this cost is $L_1 + R$. The system then returns to state 0. More than one stopping decision does not conform with the formulation of Section 3. However, this will cause no difficulty since for each λ -stopping problem we can identify the lowest cost stopping decision.

There is some loss of generality with our formulation of the state space. Once the system is in state $(i, 0)$, $0 \leq i \leq N$, in the Rosenfield notation the formulations are the same. However, if the initial state is (i, k) , $k > 0$, then our policies do not apply until the first inspection or repair takes place.

In order to apply the Regenerative Stopping Algorithm, we check Assumptions 1-3 for the λ -stopping problem in order to determine Λ . For $\lambda < L_N$ let $S = \{N\}$. It is easily seen that Assumption 3 holds. The assumption that $\sup_{i \in S^c} (\max_{a \in A_1} C(i, a))$ is bounded does not hold since a can be arbitrarily large. However, we will see that only actions satisfying $\sum_{j=0}^N P_{1j}^a L_j \leq \lambda \leq L_N$ need be considered, and for this finite set of actions the assumption in question does hold. Assumption 11 is satisfied by the stopping decision 0-repair whose cost is $R - \lambda$. Thus $\Lambda = (-\infty, L_N)$. It will not necessarily be the case that there is a $\bar{\lambda} \in \Lambda$ such that $V(\bar{\lambda}) < 0$, and for certain parameters it is optimal never to repair. For these problems the Regenerative Stopping Algorithm does not apply.

For any λ -stopping problem the set $B = \{i: C(i, a) \leq \min_{a \in A_1} [C(i, a) + \sum_j P_{1j}^a C(j, a)]\}$ where $C(j, a)$ is the cost incurred using the minimum cost stopping decision when in state j for the λ -stopping problem. The $C(i, a)$ are the continue costs for the λ -stopping problem. The set B is rather difficult to determine, so instead we consider $D = \{i: L_1 > \lambda\}$. This set is closed and is a subset of B since if $L_1 > \lambda$ then $C(i, a) > 0$ and $\sum_j P_{1j}^a C(j, a) \geq C(i, a)$ by

Rosenfield's lemma as the $C(j,s)$ are increasing in j so that $i \in B$.

As a preliminary to solving the λ -stopping problem we determine the minimum cost stopping decisions $C(i,s)$. Let $\alpha(i)$ be the largest a , $a \geq 0$, such that $\sum_j P_{ij}^a L_j < \lambda$. If $L_i > \lambda$ then no a satisfies the previous inequality and by convention we set $\alpha(i) = -1$. With this convention the minimum cost stopping decision for state i is $\alpha(i) + 1$ -repair. The cost incurred using this decision, $C(i,s)$, is $(L_i + \dots + \sum_j P_{ij}^{\alpha(i)} L_j + R - (\alpha(i) + 2)\lambda)$. If $\alpha(i) = -1$, then $C(i,s) = R - \lambda$.

The λ -stopping problem is solved as follows. For $i \in D = \{i: L_i > \lambda\}$, Theorem 1 applies and we stop. For these states $\alpha(i) = -1$, and $G(i,\lambda) = C(i,s) = R - \lambda$. For $i \notin D$ we compute G by the standard equation of optimality starting with the largest state and going down to 0. Thus we have

$$G(i,\lambda) = \min (C(i,s), \min_{0 \leq a \leq \alpha(i)} (C(i,a) + \sum_j P_{ij}^{a+1} G(j,\lambda))). \quad (8)$$

The $C(i,a)$ in (8) equals $(L_i + \sum_j P_{ij}^a L_j + \dots + \sum_j P_{ij}^a L_j + M - (a+1)\lambda)$. The states j such that $P_{ij}^a > 0$ are necessarily larger or equal to i , and G has previously been evaluated for those states. We only need to consider action a -inspect such that $a \leq \alpha(i)$, since $C(i,a)$ is increasing in a for $a \geq \alpha(i)$ and by Rosenfield's lemma, $\sum_j P_{ij}^a G(j,\lambda)$ is increasing in a .

The initial lower bound λ_0 on the optimal average cost is obtained by assuming that an inspection brings the system back to state 0, so that our costs are as with an inspection, but we get the benefit of a repair. In this case

$$\lambda_0 = \min_{a \geq 0} (L_0 + \dots + \sum_j P_{0j}^{a-1} L_j + M) / (a+1). \quad (9)$$

Example 3. We assume that there are 3 states 0, 1, and 2. The cost R of repair is 40, and the cost M of inspection is 5; $L_0 = 0$, $L_1 = 10$, and $L_2 = 20$.

The transition matrix is

$$P = \begin{bmatrix} .8 & .1 & .1 \\ 0 & .9 & .1 \\ 0 & 0 & 1.0 \end{bmatrix}$$

We will not try to determine a $\bar{\lambda} \in \Lambda$ such that $V(\bar{\lambda}) < 0$ and will address this problem at step O.B. of the Regenerative Stopping Algorithm. We calculate λ_0 by (9). The minimizing a is 1 and $\lambda_0 = (0+5)/2 = 2.5$. The 2.5-stopping is solved. The set $D = \{1,2\}$ so that $G(2,2.5) = G(1,2.5) = 40 - 2.5 = 37.5$. For state 0 we determine $\alpha(0)$ which is 0 since $0 < 2.5$ and $[\cdot 8, \cdot 1, \cdot 1]$ times $[0,10,20] = 3 > 2.5$. Therefore $G(0,2.5) = \min (40 - 5, 0 + 5 - 2.5 + .8 G(0,2.5) + .2(37.5)) = 35 = V(2.5)$. The first term in the parentheses is the cost of the decision 1-repair and the second term is the cost of the decision 0-inspect. The minimum is achieved with the decision 1-repair and the average cost period is $2.5 + 35/2 = 20$. This presents a difficulty since $20 \notin \Lambda$ and cannot be used. We arbitrarily choose a $\lambda \in \Lambda$ and hope that $V(\lambda) < 0$. If $V(\lambda) < 0$ does not obtain, we will try a larger $\lambda \in \Lambda$. This arbitrarily chosen λ is 17.5 and we solve the 17.5-stopping problem. $D = \{2\}$ so that $G(2,17.5) = 40 - 17.5 = 22.5$. For state 1 $\alpha(1)$ equals 13. The minimizing action for state 1 is to inspect after 5 periods. For state 0 $\alpha(0)$ equals 14. The minimizing action for state 0 is to inspect after 5 periods. $V(17.5) = G(0,17.5) = -84.816$ and the expected time until stopping is 12.805. Thus 17.5 can play the role of $\bar{\lambda}$.

The next λ used was a weighted average of $.9 \lambda_A$ and $.1 \lambda_B$ and equaled 9.481. This was lower than the average cost of the λ_1 -stopping problem, $17.5 - (84.816/12.805)$. This stopping problem is solved and $V(9.481) = G(0,9.481) = -.107$.

The next λ used is 9.448. The stopping problem is solved and $V(9.448) = G(0,9.448) = .1327$. The optimal policies for the 9.481 and 9.448 stopping problems are the same so that Proposition 2 can be used to confirm optimality.

This policy is 0-repair for states 1 and 2 and 2-inspect for state 0. For $\lambda = 9.448$, $G(2,9.448) = G(1,9.448) = 30.552$. For state 0 $\alpha(0) = 4$ and the optimal decision is 2-inspect. The value of $G(0,9.448) = .1327$ was obtained by solving $G(0,9.448) = -19.8445 + 5 + .488 (30.552) + .512 G(0,9.448)$.

REFERENCES

1. Abel-Hameed, M., "Optimality of the One Step Look-Ahead Stopping Times," J. of Applied Probability, 14, 162-169, 1977.
2. Bell, C., "Improved Algorithms for Inventory and Replacement-Stocking Problems," SIAM. J. of Applied Math., 18, pp. 558-566, 1970.
3. Breiman, L., "Stopping-Rule Problems," Chapter 10 of Applied Combinatorial Mathematics, E. Beckenbach, ed., Wiley, New York, 1964.
4. Brender, D., "A Surveillance Model for Recurrent Events," IBM Watson Research Center Report, Yorktown Heights, New York, 1963. Also summarized in Barlow, R. and F. Proschan, Mathematical Theory of Reliability, Wiley, New York, 1965, pp. 115-116.
5. Buckman, A. G., and B. Miller, "Optimal Investigation as a Regenerative Stopping Problem," Western Management Science Institute, UCLA, Working Paper No. 289, 1979.
6. Chow, Y., H. Robbins, and D. Siegmund, Great Expectations: The Theory of Optimal Stopping, Houghton Mifflin, Boston, 1971.
7. Denardo, E., Contraction Mappings in the Theory Underlying Dynamic Programming, SIAM Review, 9, pp. 165-177, 1967.
8. Denardo, E., "Markov Renewal Programming with Small Interest Rates," Annals of Math. Statist., 42, pp. 477-496, 1971.
9. Derman, C., and G. Lieberman, "A Markovian Decision Model for a Joint Replacement and Stocking Problem," Management Science, 13, pp. 609-617, 1967.
10. Derman, C. and J. Sacks, "Replacement of Periodically Inspected Equipment," Naval Research Logistics Quarterly, 7, pp. 597-607, 1960.
11. Feldman, R., "Optimal Replacement with Semi-Markov Shock Models," J. of Applied Probability, 13, pp. 108-117, 1976.
12. Fox, B., "The Existence of Pure Stational Optimal Policies for a Class of Markov Renewal Programs," SIAM Review, 9, pp. 573-576, 1967.
13. Jewell, W., "Markov Renewal Programming I and II," Operations Research, 11, pp. 938-971, 1963.
14. Kaplan, R., "Optimal Investigation Strategies with Imperfect Information," J. of Accounting Research, 7, pp. 32-43, 1969.
15. Lippman, S., "Maximal Average-reward Policies for Semi-Markov Decision Processes with Arbitrary State and Action Space," Annals of Math. Statist., 42, pp. 1717-1726, 1971.

16. Pierskalla, W., and J. Voelker, "A Survey of Maintenance Models: The Control and Surveillance of Deteriorating Systems," Naval Research Logistics Quarterly, 23, pp. 353-388, 1976.
17. Puterman, M., and S. Brumelle, "On the Convergence of Policy Iteration in Stationary Dynamic Programming," Mathematics of Operations Research, 4, 60-69, 1979.
18. Rosenfield, D., "Markovian Deterioration with Uncertain Information," Operations Research, 24, pp. 141-155, 1976.
19. Ross, S., "Average Cost Semi-Markov Decision Processes," J. of Applied Probability, 7, pp. 649-656, 1970.
20. Ross, S., Applied Probability Models with Optimization Applications, Holden-Day, San Francisco, 1970.
21. Taylor, H., "Optimal Replacement Under Additive Damage and Other Failure Models," Naval Research Logistics Quarterly, 22, pp. 1-18, 1975.