The Annals of Probability
1977, Vol. 5, No. 3, 413–429

OPTIMAL REPLACEMENT FOR SYSTEMS GOVERNED BY MARKOV ADDITIVE SHOCK PROCESSES

BY RICHARD M. FELDMAN

Texas A & M University

Consider a system subject to periods of deterioration. The system might fail at any time within the set of deterioration time, and the probability of failure is a function of the accumulated damage caused from past deterioration. When the system fails, it is immediately replaced and a failure cost is incurred; if replacement is made before failure, a lesser cost is incurred and that cost may depend upon the amount of accumulated damage at the replacement time.

The purpose of this paper is to derive the optimal replacement policy for such a system whose set of deterioration times contains no isolated points and whose cumulative damage process is a semi-Markov process. Only those policies which make a replacement within the set of deterioration times are considered. Optimality is based on a discounted cost criterion.

1. Introduction. Consider a system subject to failure. Let the failure depend on the amount of accumulated damage caused from past shocks, the collection of shock times being a random set. An optimal replacement problem will be analyzed in this paper under the conditions that the cumulative damage process is semi-Markovian. Optimality is based on a discounted cost criterion where a cost is incurred at replacement time.

In the situation where there are (almost surely) a finite number of shocks in any finite-length time interval, both an average cost and a discounted cost optimal replacement problem have been solved (see Feldman (1976a) and (1976b)). The situation where shocks may occur continuously during a time interval, or more generally, where the collection of shock times forms a random set containing no isolated points, is discussed here.

If the set of regeneration points (shock times) for a semi-Markov process contains no isolated points almost surely, then the semi-Markov process (under fairly loose conditions) contains an imbedded Markov additive process. In this paper, the imbedded Markov additive process is used to solve the optimal replacement problem. In Section 2, a brief review of the relevant definitions for semi-Markov and Markov additive processes is given and the optimal replacement problem is explicitly stated. In Section 3, the theory for the optimal stopping of a Markov additive process is developed. In Section 4, the optimal

Received March 2, 1976; revised July 19, 1976.

[1] Research partially supported by the Air Force Office of Scientific Research through their Grant Number AFOSR-74-2733. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

AMS 1970 *subject classifications.* Primary 90B25, 60K99; Secondary 60G40, 60J25.

Key words and phrases. Optimal replacement, Markov additive processes, semi-Markov processes, optimal stopping theory.

stopping theory is used to obtain the optimal replacement time. In Section 5, some computational aspects are given.

To illustrate the use of Markov additive processes, consider a machine subject to deterioration. Let $f(x)$ denote the cost of replacing the machine if its deterioration level at the time of replacement is x. Let $(X_t)_{t \geqq 0}$ be a Markov process, where X_t denotes the level of deterioration if the machine is used *continuously* during the time interval $[0, t]$. However, in actual operation, the machine is not in continuous use. The length of intervals of operating time follows an exponential distribution and the length of intervals of down time follows an arbitrary distribution that possibly depends on the level of deterioration. Let $(C_t)_{t \geqq 0}$ be the "clock time" process; that is, C_t denotes the actual amount of time that the machine was used during $[0, t]$. Let Y be the inverse of the clock time C; that is

$$(1.1) \qquad Y_t = \inf\{s \geqq 0 : C_s > t\}\,.$$

The real time damage process is given by $(Z_t)_{t \geqq 0}$ which is defined by

$$(1.2) \qquad \begin{aligned} Z_t &= X_{C_t} && \text{if} \quad Y_s = C_t \quad \text{for some} \quad s\,, \\ &= X_{C_t -} && \text{otherwise.} \end{aligned}$$

The real time process Z is a semi-Markov process and (X, Y) is a Markov additive process. (It might seem more natural to define Z by setting $Z_t = X_{C_t}$ for each t. The definition of (1.2) is used instead so that the *initiation* of an operating interval may cause a jump in the deterioration level.)

We close this section by giving some of the notations and conventions to be used throughout the paper. In general, the notations follow Blumenthal and Getoor (1968). The letter E will denote the state space of a stochastic process and can be taken as an interval of $\mathbb{R}$ (real) or of $\mathbb{N}$ (nonnegative integers). The set E will be closed on the left and the letter Δ will designate an element adjoined to E such that $\Delta > x$ for all $x \in E$. The point $\omega_\Delta \in \Omega$ is such that for a process $Z = (Z_t)_{t \geqq 0}$, we have $Z_0(\omega_\Delta) = \Delta$.

Using the terminology of Dellacherie (1972), a random set G of the probability space $(\Omega, \mathscr{F}, P)$ is a mapping from Ω into $\mathscr{R}_+$, i.e., for each $\omega \in \Omega$, $G(\omega)$ is a Borel subset of $\mathbb{R}_+$. The random set G is called right closed if, for every $\omega \in \Omega$, $G(\omega)$ contains the limit of any decreasing sequence of points contained in $G(\omega)$. The contiguous intervals of G are the connected components of $\mathbb{R}_+ \backslash G$. A random set G is called minimal if the left extreme points of its contiguous intervals are not in G almost surely. (The term minimal means it is the smallest right closed set having the given closure.) The random set G is called perfect if for almost every ω, $G(\omega)$ contains no isolated points. If T is a random variable and G is a random set such that $T(\omega) \in G(\omega)$ for almost all $\omega \in \{T < \infty\}$, T is said to be contained in G and is written $T \subset G$. For this paper, a random set G is always assumed to contain zero and be minimal, perfect and right closed.

2. Preliminary difinitions and problem formulation. The process Z is semi-Markovian if, loosely speaking, for any stopping time $T \subset G$ the future of Z is independent of the past given in the current value Z_T. The semi-Markov process Z is constant over the contiguous intervals of G, and for $t \notin G(\omega)$, the value of $Z_t(\omega)$ is equal to $\lim_{s \to U_t -} Z_s(\omega)$ where $U_t(\omega) = \sup\{s \leqq t: s \in G(\omega)\}$; that is, Z_t is determined by the left closest portion of G. The definition of a semi-Markov process used in this paper is due to Jacod (1974). The value of the semi-Markov process over the contiguous intervals is the main difference between the definition of Jacod and the definition of Maisonneuve (1974). Maisonneuve assumes the process can jump at the left endpoints, but is continuous at the right endpoint of the contiguous intervals. Jacod assumes the process is continuous at the left endpoints and can jump at the right endpoints of the contiguous intervals.

The formal definitions for the processes to be used throughout this paper will now be given.

(2.1) DEFINITION. The process $Z = (\Omega, \mathscr{F}, \mathscr{F}_t, Z_t, \theta_t, P^z; G)$ is called a standard semi-Markov process with state space $(E, \mathscr{E})$ and lifetime $\tau = \inf\{t \geqq 0: Z_t = \Delta\}$ if the following properties hold:

(a) G is a right-closed, random set, includes 0 if $\omega \neq \omega_\Delta$, and $G(\omega_\Delta) = \emptyset$.

(b) The mapping $\omega \to 1_{G(\omega)}(t)$ is in $\mathscr{F}_t$.

(c) Z is adapted to $(\mathscr{F}_t)_{t \geqq 0}$, is right-continuous, has left-hand limits, $Z_t(\omega) = \Delta$ for $t \geqq \tau(\omega)$, and $Z_t \circ \theta_s = Z_{t+s}$ for $t, s \geqq 0$.

(d) $\lim_{n\to\infty} Z_{T_n} = Z_T$ on $\{T < \tau\} P^z$-a.s. for any $z \in E$ if $T_n \to T$ where $(T_n)_{n \in \mathbf{N}}$ is an increasing sequence of stopping times contained in G.

(e) $P^z\{Z_0 = z\} = 1$ for all $z \in E$.

(f) $E^z[f(Z_{T+t}) \mid \mathscr{F}_T] = E^{Z(T)}[f(Z_t)]$ for any $z \in E$, $f \in b\mathscr{E}_\Delta$, and stopping time $T \subset G$.

(g) $t \to Z_t(\omega)$ is constant over contiguous intervals of G and is left-continuous at each $t \notin G(\omega)$.

(h) Any $(\mathscr{F}_t)$ stopping time $T \subset \bar{G} - G$ is totally inaccessible on $\{T < \tau\}$.

The set G is called a semiregenerative set and (Z, G) is called a semiregenerative system.

(2.2) REMARK. By conditions (a) and (b), we have that the process $(1_G(t))_{t \geqq 0}$ is progressively measurable. The semi-Markov process Z will induce a Markov process; conditions (c), (d) and (e) will be used to insure that the induced Markov process to be defined below is a standard Markov process. Condition (f) is the key semiregenerative property. Condition (g) indicates that the value of Z over the contiguous intervals is determined by the value of Z in G on the left. Condition (h) is more easily understood when considered in conjunction with the induced Markov additive process and thus will be discussed later (see the remark after Proposition (2.10)). It will turn out that the condition (h) is not very restrictive.

Associated with the semiregenerative system (Z, G) is an additive functional

$(L_t)_{t\geqq 0}$ of Z such that the set of points of increase of (L_t) is indistinguishable from $\bar{G}$. Such an additive functional is called a local time of (Z, G) and plays a key role in defining the imbedded Markov additive process.

(2.3) PROPOSITION. *Let $Z = (\Omega, \mathscr{F}, \mathscr{F}_t, Z_t, \theta_t, P^z; G)$ be a standard semi-Markov process and let G be minimal and perfect. There exists a unique continuous increasing process $(L_t)_{t\geqq 0}$ adapted to $(\mathscr{F}_t)_{t\geqq 0}$ with the following properties*:

(a) $L_0(\omega) = 0$ *for* $\omega \in \Omega$.

(b) $L_{t+s}(\omega) = L_t(\omega) + L_s \circ \theta_t(\omega)$ *for all* $s, t > 0$ *and* $\omega \in \Omega'$ *where* $\Omega' = \Omega - \Lambda$ *for some* Λ *with* $P^z(\Lambda) = 0$ *for all* $z \in E$.

(c) *The set of points of increase of* (L_t) *is indistinguishable from* $\bar{G}$ *and the set of points of right increase of* (L_t) *is indistinguishable from* G;

(d) $E^z[\int_0^\infty e^{-t}L(dt)] = 1$ *for* $z \in E$.

PROOF. This proof is given in Maisonneuve ((1974), page 66). Condition (b) implies that L is a perfect additive functional and condition (d) gives the uniqueness. It should be noted that condition (h) of Definition (2.1) is needed for L to be continuous. □

As an example of the local time, the process $C = (C_t)_{t\geqq 0}$ from Section 1 is a local time for (Z, G). In this case Z is as defined by (1.2) and $G(\omega)$ is the range of the function $t \to Y_t(\omega)$ defined by (1.1).

Let $\hat{Y}$ designate the time inverse of L, let $\hat{X}$ denote the time changed process of Z, and let $\hat{\zeta}$ denote the maximum value of L.

$$\hat{Y}_t(\omega) = \inf\{s \geqq 0 : L_s(\omega) > t\} \qquad \text{for} \quad t \geqq 0, \quad \omega \in \Omega;$$
$$\hat{X}_t(\omega) = Z(\hat{Y}_t(\omega), \omega) \qquad \text{for} \quad t \geqq 0, \quad \omega \in \Omega;$$
$$\hat{\zeta}_t(\omega) = L_\infty(\omega) = \inf\{t \geqq 0 : \hat{Y}_t(\omega) = \infty\} \qquad \text{for} \quad \omega \in \Omega.$$

We now form the process (X, Y). Define ζ, Y, X and θ_t' for each $\omega \in \Omega$ by

$$\zeta(\omega) = \hat{\zeta}(\omega) \wedge \inf\{t \geqq 0 : \hat{X}_t(\omega) = \Delta\}; \tag{2.4}$$

$$\begin{aligned} Y_t(\omega) &= \hat{Y}_t(\omega) && \text{if} \quad t < \zeta(\omega), \\ &= \hat{Y}_\zeta(\omega) && \text{if} \quad t \geqq \zeta(\omega); \end{aligned} \tag{2.5}$$

$$\begin{aligned} X_t(\omega) &= \hat{X}_t(\omega) && \text{if} \quad t < \zeta(\omega), \\ &= \Delta && \text{if} \quad t \geqq \zeta(\omega); \end{aligned} \tag{2.6}$$

$$\begin{aligned} \theta_t' &= \theta(\hat{Y}_t(\omega), \omega) && \text{if} \quad t < \zeta(\omega), \\ &= \omega_\Delta && \text{if} \quad t \geqq \zeta(\omega). \end{aligned} \tag{2.7}$$

Let $(\mathscr{M}_t)_{t\geqq 0}$ be the canonical family of σ-algebras generated by (X, Y). Let $\mathscr{M} = \mathscr{M}_\infty$. Define P'^z in the obvious manner.

The result we are leading up to is the theorem of Jacod (1974) stating that (X, Y) is a Markov additive process. Before giving Jacod's theorem, we first give the definition of a Markov additive process from Çinlar (1972).

(2.8) DEFINITION. The process $(X, Y) = (\Omega, \mathscr{M}, \mathscr{M}_t, X_t, Y_t, \theta_t, P^x)$ is called a Markov additive process with state space $(E, \mathscr{E})$ if the following hold:

(a) $X = (\Omega, \mathscr{M}, \mathscr{M}_t, X_t, \theta_t, P^x)$ is a standard Markov process with state space $(E, \mathscr{E})$ (in the sense of Blumenthal and Getoor (1968), page 45);

(b) for almost all $\omega \in \Omega$, $t \to Y_t(\omega)$ is right continuous, has left-hand limits, $Y_0(\omega) = 0$, and $Y_t(\omega) = Y_\zeta(\omega)$ for $t \geqq \zeta$ where $\zeta(\omega) = \inf\{t \geqq 0 : X_t(\omega) = \Delta\}$;

(c) Y is adapted to $(\mathscr{M}_t)$;

(d) $x \to P^x\{X_t \in A, Y_t \in B\}$ is in $\mathscr{E}$ for each $t \geqq 0$, $A \in \mathscr{E}$, $B \in \mathscr{R}_+$;

(e) $Y_{t+s}(\omega) = Y_t(\omega) + Y_s \circ \theta_t(\omega)$ for all $s, t \geqq 0$ and $\omega \in \Omega'$ where $\Omega' = \Omega - \Lambda$ for some Λ with $P^x(\Lambda) = 0$ for all $x \in E$;

(f) $P^x\{X_s \circ \theta_t \in A, Y_s \circ \theta_t \in B \mid \mathscr{M}_t\} = P^{X(t)}\{X_s \in A, Y_s \in B\}$ for all $t, s \geqq 0$, $x \in E_\Delta$, $A \in \mathscr{E}_\Delta$, $B \in \mathscr{R}_+$.

REMARK. Definition (2.8) is more restrictive than as introduced by Çinlar (1972). Namely, we insist that X be a standard process, that the state space of Y be $(\mathbb{R}_+, \mathscr{R}_+)$, and that (X, Y) be perfect.

(2.9) THEOREM. *Let $Z = (\Omega, \mathscr{F}, \mathscr{F}_t, Z_t, \theta_t, P^z; G)$ be a standard semi-Markov process. Let G be minimal and perfect. Then the process $(X, Y) = (\Omega, \mathscr{M}, \mathscr{M}_t, X_t, Y_t, \theta_t', P'^x)$ defined by equations* (2.5) *through* (2.7) *is a Markov additive process. Furthermore, $t \to Y_t(\omega)$ is strictly increasing on $[0, \zeta]$ almost surely.*

PROOF. This theorem differs from Jacod ((1974), Proposition 2.4) only in that we say that X is a standard Markov process. The only property necessary for a standard process not obvious from Jacod's proof is quasi-left continuity. To prove this, first define the random set $\hat{N}$ by

$$\hat{N}(\omega) = \{t \geqq 0 : Y_{t-}(\omega) < Y_t(\omega)\}\ .$$

Jacod ((1974), Proposition 2.3) shows that any stopping time of $(\mathscr{M}_t)$ contained in N is totally inaccessible (condition (h) of Definition (2.1) is necessary for this); thus, if T is a predictable $(\mathscr{M}_t)$ stopping time, then $Y_{T-}(\omega) = Y_T(\omega)$. If T_n is an increasing sequence of $(\mathscr{M}_t)$ stopping times converging to a predictable stopping time T, then $Y_{T_n} \to Y_T$ a.s. The fact that $Y(T_n)$ is an $(\mathscr{F}_t)$ stopping time follows from a lemma of Jacod which is repeated in this paper as Theorem (2.14). By condition (d) of Definition (2.1), $Z(Y(T_n)) \to Z(Y(T))$ a.s. which gives $X(T_n) \to X(T)$ a.s. □

For the Markov additive process (X, Y), the process Y has been completely characterized by Çinlar (1972). We summarize the characterization of Y in the following proposition.

(2.10) PROPOSITION. *Let (X, Y) be a Markov additive process. The process Y can be decomposed as*

$$Y_t = A_t^c + Y_t^p + Y_t^q + Y_t^d \qquad \text{for} \quad t \geqq 0\ ,$$

where the components are conditionally independent given X and satisfy the following:

(a) $A^c = (A_t^c)_{t \geqq 0}$ *is a continuous additive functional of* X;

(b) $Y^p = (Y_t^p)_{t \geqq 0}$ *is a predictable pure jump process whose jump times are fixed by* X;

(c) $Y^q = (Y_t^q)_{t \geqq 0}$ *is a quasi-left continuous pure jump process whose jump times are fixed by* X;

(d) $Y^d = (Y_t^d)_{t \geqq 0}$ *is a conditionally stochastically continuous process given* X *whose jump times are not fixed by* X.

PROOF. This proposition is a restatement of results given in Çinlar ((1974), page 11) and its proof is in Çinlar ((1972), Theorems 2.23 and 4.5). □

REMARK. The predictable property and the quasi-left continuous property refer to the σ-algebras generated by X. A better intutitive feel for these components can be obtained going through the computations in Section 5. Since, for this paper, $t \to Y_t(\omega)$ is increasing, there is no Gaussian component of Y. The predictable component Y^p is sometimes called the natural component because it has almost surely no common discontinuities with $t \to X_t(\omega)$.

Condition (h) of Definition (2.1) can now be interpreted to be equivalent to insisting that $Y^p = 0$. Thus, the Markov additive process (X, Y) of Theorem (2.9) has $Y^p = 0$.

If we start with a Markov additive process (X, Y) such that $Y^p = 0$ and $t \to Y_t(\omega)$ is strictly increasing, a semiregenerative system (Z, G) with G minimal and perfect is easily obtained by using the following definitions:

(2.11) $$G(\omega) = \{y \geqq 0 : Y_t(\omega) = y \text{ for some } t \geqq 0\};$$

(2.12) $$L_t(\omega) = \inf\{s \geqq 0 : Y_s(\omega) > t\};$$

(2.13) $$\begin{aligned} Z_t(\omega) &= X_{L(t)}(\omega) && \text{if } t \in G(\omega), \\ &= X_{L(t)-}(\omega) && \text{if } t \notin G(\omega). \end{aligned}$$

Our goal is to investigate properties of the Markov additive process instead of the semi-Markov process. For this reason we repeat the relationship between the stopping times of the two processes as given by Jacod (1974).

(2.14) THEOREM. *Let* $Z = (\Omega, \mathscr{F}, \mathscr{F}_t, Z_t, \theta_t, P^z; G)$ *be a standard semi-Markov process such that* G *is minimal and perfect, let* $L = (L_t)_{t \geqq 0}$ *be the local time of* (Z, G), *and let* $(X, Y) = (\Omega, \mathscr{M}, \mathscr{M}_t, X_t, Y_t, \theta_t', P'^x)$ *be its imbedded Markov additive process. If* T *is an* $(\mathscr{F}_t)$ *stopping time, then* L_T *is an* $(\mathscr{M}_t)$ *stopping time; if* S *is an* $(\mathscr{M}_t)$ *stopping time, then* Y_S *is an* $(\mathscr{F}_t)$ *stopping time.*

PROOF. See Jacod ((1974), Lemma 2.2).

To formulate the optimal replacement problem using discounted costs, let $f(z)$ be the negative of the cost of replacing the system if at the time of replacement the system is in state z. (In order to use the existing optimal stopping theory without unnecessary technical detail, it is easier to work with a maximizing problem instead of a minimizing problem. For this reason, the payoff function will be negative.) Thus $-f(\Delta)$ is the cost of replacing a failed system. (Note that for $f \in E_\Delta$, $f(\Delta)$ need *not* be zero.) The optimal replacement problem can be stated in two equivalent formulations.

(2.15) PROBLEM. Let $Z = (\Omega, \mathscr{F}, \mathscr{F}_t, Z_t, \theta_t, P^z; G)$ be a semi-Markov process. Let G be a perfect minimal set. Let $\tau = \inf\{t \geqq 0 : Z_t = \Delta\}$ and let $\mathscr{S} = \{S \leqq \tau : S \text{ is an } (\mathscr{F}_t) \text{ stopping time and } S \subset G\}$. For fixed $\lambda > 0$ and for each $z \in E$, determine $w(z)$, where $w(z)$ is defined by

$$w(z) = \sup_{S \in \mathscr{S}} E^z[e^{-\lambda S} f(Z_S)] ,$$

and if possible, find $S_0 \in \mathscr{S}$ such that

$$E^z[e^{-\lambda S_0} f(Z_{S_0})] = w(z) .$$

(2.16) PROBLEM. Let $(X, Y) = (\Omega, \mathscr{M}, \mathscr{M}_t, X_t, Y_t, \theta_t, P^x)$ be the Markov additive procees imbedded in the semi-Markov process of Problem (1.1). Let $\zeta = \inf\{t \geqq 0 : X_t = \Delta\}$ and let $\mathscr{T} = \{T \leqq \zeta : T \text{ is an } (\mathscr{M}_t) \text{ stopping time}\}$. For a fixed $\lambda > 0$ and for each $x \in E$, determine $v(x)$, where $v(x)$ is defined by

$$v(x) = \sup_{T \in \mathscr{T}} E^x[e^{-\lambda Y(T)} f(X_T)]$$

and, if possible, find $T_0 \in \mathscr{T}$ such that

$$E^x[e^{-\lambda Y(T_0)} f(X_{T_0})] = v(x) .$$

(2.17) THEOREM. *If Y is strictly increasing, then Problem* (2.15) *and Problem* (2.16) *are equivalent.*

PROOF. By Theorem (2.14), we have for $S \in \mathscr{S}$, $L_S \in \mathscr{T}$; and for $T \in \mathscr{T}$, $Y_T \in \mathscr{S}$. Using equation (2.13) and letting $S \in \mathscr{S}$ and $T = L_S$, the functions $e^{-\lambda S} f(Z_S)$ and $e^{-\lambda Y(T)} f(X_T)$ are equal for all $\omega \in \Omega$, and thus their expectations are equal. Using equation (2.6) and letting $T \in \mathscr{T}$ and $S = Y_T$, the same is again true. To show $w(z) = v(z)$, let $\{S_n; n \in \mathbb{N}\}$ be a sequence contained in $\mathscr{S}$ such that

$$\lim_{n\to\infty} E^z[e^{-\lambda S_n} f(Z_{S_n})] = w(z) .$$

For each n, let $T_n = L_{S_n}$; then if $w(z) \neq v(z)$, we would have for some $T^* \in \mathscr{T}$

$$\lim_{n\to\infty} E^z[e^{-\lambda Y(T_n)} f(X_{T_n})] < E^z[e^{-\lambda Y(T^*)} f(X_{T^*})] .$$

By letting $S^* = Y_{T^*}$ a contradiction is obtained. □

(2.18) COROLLARY. *If T_0 is a solution to Problem* (2.16), *then Y_{T_0} is a solution to Problem* (2.15).

PROOF. Obvious from the proof of (1.3). □

3. Optimal stopping of a Markov additive process. Problem (2.16) expresses the semi-Markov optimal replacement problem as a Markov additive optimal stopping problem. The theory for the optimal stopping of a Markov process with continuous paths (see Fakeev (1971)) and a Markov process with a Feller transition function (see Taylor (1968)) has been developed. In this section, the theory for the optimal stopping of a Markov additive process will be developed. The general approach will follow that used by Fakeev (1971); however, the assumptions used will (among other things) imply a Feller-like property (see (3.3)) but not continuous paths. Fakeev assumes that both f and $t \to X_t$ are continuous (he calls X an "unbroken" process). If it were not for these assumptions, Fakeev's results could be applied directly since the Markov additive process (X, Y) is also a Markov process with state space $E \times \mathbb{R}_+$.

The results of this section are analogous to the results for Markov processes in which excessive functions play a key role. Fakeev's (1971) definition of an excessive function for a Markov process is the traditional definition except that the function is not restricted to be positive. In a similar manner, we shall define an excessive function for a Markov additive process in the traditional manner (see Maisonneuve (1974), Definition 5.4)) except without the restriction of positivity.

Let (X, Y) be a Markov additive process with $\zeta = \inf\{t \geqq 0 : X(t) = \Delta\}$. Let $\lambda > 0$ be a fixed discount rate. For any $(\mathscr{M}_t)$ stopping time T, the operator $Q_T{}^\lambda$ is defined by

$$(3.1) \qquad Q_T{}^\lambda g(x) = E^x[e^{-\lambda Y(T)} g(X_T)] \qquad \text{for} \quad g \in \mathscr{E}_\Delta .$$

(3.2) Definition. The function $g \in \mathscr{E}_\Delta$ is called λ-excessive if the following two properties hold:

(a) $\lim_{t \to 0} Q_t{}^\lambda g(x) = g(x)$ for $x \in E$,
(b) $g \geqq Q_t{}^\lambda g$ for each $t \geqq 0$.

The following assumptions will be used throughout the remainder of this paper.

(3.3) Assumptions. Let (X, Y) be a Markov additive process and let $f \in b\mathscr{E}_\Delta$ be the payoff function, then

(a) X is a Hunt process and $t \to X_t(\omega)$ is nondecreasing for all $\omega \in \Omega$;
(b) $t \to Y_t(\omega)$ is strictly increasing for all $\omega \in \Omega$;
(c) f is nonpositive, nonincreasing, and right-continuous on E_Δ;
(d) $x \to P^x\{X_t \geqq r, Y_t < y\}$ is nondecreasing and right-continuous on E_Δ for $r \in E$ and $t, y \geqq 0$.

Remarks. Condition (a) implies that the system cannot repair itself and X quasi left-continuous on $[0, \infty)$. Conditiion (b) is due to the assumption that the semiregenerative set representing times of deterioration is minimal and perfect. Condition (c) indicates that costs instead of profits are involved, that the cost increases as damage increases, and that the process $(f(X_t))_{t \geqq 0}$ is right-continuous. Condition (d) implies that as the initial state gets worse, the damage stochastically increases, while the time actual damage occurs stochastically decreases. Condition (d) also implies a right-continuous version of the Feller property. One of the main uses of Assumption (3.3d) is the following lemma. (See Serfozo (1977), Section 4) for general results similar to Lemma (3.4).)

(3.4) LEMMA. *For each $x \in E$, let $(\Omega, \mathscr{F}, P^x)$ be a probability space and let X and T be two random variables with their range in E and $\mathbb{R}_+$ respectively. Let $g \in b\mathscr{E}_\Delta$ and $h \in b\mathbb{R}_+$ satisfy the following*:

(a) *g is nonpositive, nonincreasing and right-continuous*;
(b) *h is nonnegative, nonincreasing and right-continuous.*

If $x \to P^x\{X \geqq y, T < t\}$ is nondecreasing and right-continuous for each $y \in E$, $t \geqq 0$, then $x \to E^x[h(T)g(X)]$ is nonincreasing and right-continuous.

PROOF. In this proof we make use of the fact that a decreasing right-continuous function is lower semicontinuous. Several proofs in this section use the elementary properties of lower semicontinuous functions as found in Royden ((1968), pages 48–49).

Assume first that g and h are right-continuous decreasing simple functions. They can then be written as

$$g(x) = \sum_{n=0}^N c_n 1_{[a_n,\infty)}(x) \qquad \text{for} \quad x \in E\,,$$

where each $c_n \leqq 0$ and $a_0 < a_1 < \cdots < a_N$, and

$$h(t) = \sum_{n=0}^N d_n 1_{[0,b_n)}(t) \qquad \text{for} \quad t \in \mathbb{R}_+\,,$$

where each $d_n \geqq 0$ and $b_0 < b_1 < \cdots < b_N$. Then

$$E^x[h(T)g(X)] = \sum_{n=0}^N \sum_{k=0}^N c_n d_k P^x\{X \geqq a_n, T < b_k\}\,.$$

Since $c_n d_k \leqq 0$ and there are only finitely many terms, $x \to E^x[h(T)g(X)]$ is decreasing and right-continuous.

General functions g and h can be expressed as the limit of an increasing sequence of step functions, each step function being decreasing and right-continuous and thus, lower semicontinuous. With the monotone convergence theorem and since the supremum of lower semicontinuous functions is again lower semicontinuous, the proof is complete. □

In this section, we shall rely heavily on the results of Fakeev (1970). Define the following:

$$(3.5) \qquad \mathscr{T}_t = \{T\colon T \text{ is an } (\mathscr{M}_t) \text{ stopping time and } t \leqq T < \infty\}\,,$$

$$(3.6) \qquad W_t = e^{-\lambda Y(t)} f(X_t) \qquad \text{for} \quad t \geqq 0\,,$$

$$(3.7) \qquad v_t(x) = \sup_{T \in \mathscr{T}_t} E^x[e^{-\lambda Y(T)} f(X_T)] \qquad \text{for} \quad t \geqq 0\,, \quad x \in E_\Delta\,,$$

$$(3.8) \qquad V_t(x) = \text{ess.}\sup_{T \in \mathscr{T}_t} E^x[e^{-\lambda Y(T)} f(X_T) \mid \mathscr{M}_t] \qquad \text{for} \quad t \geqq 0\,, \quad x \in E_\Delta\,.$$

By Fakeev ((1970), Theorem 2), $(V_t, \mathscr{M}_t)_{t \geqq 0}$ is the minimal right-continuous supermartingale majorizing $(W_t)_{t \geqq 0}$. This fact will be utilized to show that v_0 is the minimal λ-excessive function majorizing f. (This property of v_0 is needed in the next section to show that a "control limit" type policy is the optimal policy.) In order to prove the characterization of v_0, we define a sequence of random variables $\{H_n(x, t); n \in \mathbb{N}\}$ and a sequence of functions $\{g_n; n \in \mathbb{N}\}$. First let Q_t denote the set of rational numbers greater than t. Now define the following:

$$H_0(x, t) = W_t \quad \text{for} \quad t \geqq 0, \quad x \in E_\Delta,$$

(3.9)
$$H_n(x, t) = \sup_{s \in Q_t} E^x[H_{n-1}(x, s) \mid \mathscr{M}_t] \quad \text{for} \quad t \geqq 0, \quad x \in E_\Delta, \quad n = 1, 2, \cdots,$$

(3.10)
$$g_0(x) = f(x) \quad \text{for} \quad x \in E_\Delta$$
$$g_n(x) = \sup_{s \geqq 0} E^x[e^{-\lambda Y(s)} g_{n-1}(X_s)] \quad \text{for} \quad x \in E_\Delta, \quad n \in \mathbb{N}_+ .$$

(3.11) LEMMA. *The function $t \to g_n(X_t)$ is lower semicontinuous for $n \in \mathbb{N}$.*

PROOF. For a decreasing function, right-continuity is equivalent to lower semicontinuity. If it can be shown that $x \to g_n(x)$ is decreasing and right-continuous, then by the increasing and right-continuous property of $t \to X_t$, the proof would be complete. By Assumption (3.3 c), the property holds for g_0. By induction, assume g_n is decreasing and right-continuous for a fixed n. From Lemma (3.4), we have that $x \to E^x[e^{-\lambda Y(s)} g_n(X_s)]$ is decreasing and right-continuous for each s. Since the supremum of lower semicontinuous functions is again lower semicontinuous, the proof is complete. □

(3.12) LEMMA. *For each $n \in \mathbb{N}$ and $t \geqq 0$, there is a version of the conditional expectation in* (3.9) *such that*

$$H_n(x, t) = e^{-\lambda Y(t)} g_n(X_t) \qquad \textit{for each} \quad \omega \in \Omega .$$

PROOF. The proof is by induction. Clearly the lemma is true for $n = 0$. Assume it is true for $n \geqq 1$. Then, for each $\omega \in \Omega$,

$$\begin{aligned} H_{n+1}(x, t) &= \sup_{s \in Q_t} E^x[H_n(x, s) \mid \mathscr{M}_t] \\ &= \sup_{s \in Q_0} E^x[e^{-\lambda Y(t+s)} g_n(X_{t+s}) \mid \mathscr{M}_t] \\ &= \sup_{s \in Q_0} e^{-\lambda Y(t)} E^{X(t)}[e^{-\lambda Y(s)} v_n(X_s)] \\ &= e^{-\lambda Y(t)} g_{n+1}(X_t) . \end{aligned}$$

The third equality is the Markov property as given in Çinlar ((1972), page 106). The last equality follows, since Lemma (3.11) gives right-continuity so that supremum over $s \in Q_0$ is equivalent to $s \geqq 0$. □

(3.13) LEMMA. *Let $\{H_n(x, t); n \in \mathbb{N}\}$ be defined by* (3.9) *and* (3.12), *then*

$$V_t(x) = \lim_{n \to \infty} H_n(x, t) .$$

PROOF. Using (3.10) and (3.12), we have that $\{H_n(x, t); n \in \mathbb{N}\}$ is an increasing sequence, and thus, a limit exists; call it $H(x, t)$. Let $(U_t, \mathscr{M}_t)_{t \geqq 0}$ be any supermartingale majorizing $(W_t)_{t \geqq 0}$. Clearly $U_t \geqq H_0(x, t)$. By induction, it follows that $H_{n+1}(x, t) \leqq \sup_{s \in Q_t} E^x[U_s \mid \mathscr{M}_t] \leqq U_t$, and thus $H(x, t) \leqq U_t$. By the minimality of $(V_t(x), \mathscr{M}_t)_{t \geqq 0}$, the proof will be complete when we show that $(H(x, t), \mathscr{M}_t)_{t \geqq 0}$, is a right-continuous supermartingale majorizing $(W_t)_{t \geqq 0}$.

Clearly $H(x, t) \geqq W_t$ and by the Lebesgue convergence theorem, we can pass to the limit as $n \to \infty$ in the inequality

$$H_n(x, t) \geqq E^x[H_{n-1}(x, t+s) \mid \mathscr{M}_t],$$

and obtain the supermartingale inequality. By Lemma (3.12), $H(x, t)$ is right-continuous if $t \to g(X_t) = \lim_{n\to\infty} g_n(X_t)$ is right-continuous since $t \to e^{-\lambda Y(t)}$ is right-continuous. The sequence $\{g_n; n \in \mathbb{N}\}$ is increasing, so g exists. By Lemma (3.11), $t \to g_n(X_t)$ is lower semicontinuous, and therefore, $t \to g(X_t)$ is also. Since $t \to g_n(x_t)$ is decreasing, $t \to g(X_t)$ is decreasing and right-continuity follows. □

(3.14) THEOREM. *For each* $t \geqq 0$ *and* $x \in E$, $V_t(x) = e^{-\lambda Y(t)} v_0(X_t)$.

PROOF. As above, let $g = \lim g_n$. Using Lemmas (3.13) and (3.12), we have $V_t(x) = e^{-\lambda Y(t)} g(X_t)$. By setting $t = 0$ and taking expectations on both sides, we obtain $v_0 = g$ since $E^x[V_0(x)] = v_0(x)$. □

(3.15) THEOREM. *The optimal function* v_0 *defined by equation* (3.7) *is the minimal* λ*-excessive function majorizing the payoff function* f.

PROOF. Using (3.1) and (3.14), we have $Q_t^\lambda v_0(x) = E^x[V_t(x)]$ and since $(V_t(x))_{t\geqq 0}$ is a supermartingale, v_0 is λ-excessive. To show v_0 is the minimal λ-excessive function majorizing f, let $u \in b\mathscr{E}_\Delta$ be any λ-excessive function such that $u \geqq f$. Define the process $(U_t)_{t\geqq 0}$ by

$$U_t = e^{-\lambda Y(t)} u(X_t) \qquad \text{for} \quad t \geqq 0 .$$

Clearly $t \to U_t$ is right-continuous and $U_t \geqq W_t$. By using the Markov property of a Markov additive process (Çinlar (1972), Proposition 3.15), it is easily seen that $(U_t, \mathscr{M}_t)_{t\geqq 0}$ is a supermartingale. By the minimality of $(V_t)_{t\geqq 0}$, we have $V_t(x) \leqq U_t$ for $t \geqq 0$, and in particular, $V_0(x) \leqq U_0$. Thus $E^x[V_0(x)] \leqq E^x[U_0]$, which implies $v_0(x) \leqq u(x)$. □

Theorem (3.14) gives a characterization of the supermartingale $(V_t(x))_{t\geqq 0}$. This will be combined in the next theorem with Fakeev's results to obtain the optimal stopping time. Theorem (3.15) characterizes the optimal function v_0. The properties of v_0 will be used to further characterize the optimal stopping time.

(3.16) THEOREM. *Let* f *be continuous on* E *(with possibly a jump at* Δ*). Let* $\zeta = \inf\{t \geqq 0 : X_t = \Delta\}$*. Define the random variable* T *by*

$$T = \inf\{t \geqq 0 : f(X_t) = v_0(X_t)\} .$$

Then T *is the optimal stopping time; that is*

$$v_0(x) = E^x[e^{-\lambda Y(T)} f(X_T)] .$$

PROOF. This is almost proved in Fakeev ((1970), Theorem 4). Fakeev uses continuity of W, but in fact only needs quasi left-continuity. In order to see this define, for each $\varepsilon > 0$, the random variable $S(\varepsilon)$ by

(3.17) $$S(\varepsilon) = \inf\{s \geqq 0 : e^{-\lambda Y(s)} v_0(X_s) \leqq e^{-\lambda(Y)} f(X_s) + \varepsilon\} .$$

Then by Fakeev ((1970), Theorem 4a) and using Theorem (3.14), we have

(3.18) $$v_0(x) - \varepsilon \leqq E^x[e^{-\lambda Y(S(\varepsilon))} f(X_{S(\varepsilon)})] .$$

As $\varepsilon \to 0$, $S(\varepsilon)$ increases and thus, limits to a stopping time, call it S. On the set $\{S = \zeta\}$, there must exist some $\varepsilon > 0$ (possibly depending on ω and excluding a null set) such that $S(\varepsilon) = \zeta$ since ζ cannot be predictable for a Hunt process. Therefore,

$$\lim_{\varepsilon \to \infty} e^{-\lambda Y(S(\varepsilon))} f(X_{S(\varepsilon)}) = e^{-\lambda Y(\zeta)} f(\Delta) \qquad \text{on} \quad \{S = \zeta\} .$$

On the set $\{S < \zeta\}$, the above limit is combined with the quasi left-continuous property of a standard Markov process, the fact that f is continuous on E, and the fact that the predictable part of Y is zero (see the remark after (2.10)) to obtain

(3.19) $$\lim_{\varepsilon \to 0} e^{-\lambda Y(S(\varepsilon))} f(X_{S(\varepsilon)}) = e^{-\lambda Y(S)} f(X_S) \quad \text{a.s.}$$

With this limit, the remainder of Fakeev's proof follows. We repeat it here for completeness.

Applying (3.19) to (3.18) and using Fatou's lemma,

$$\begin{aligned} v_0 &\leqq \lim_{\varepsilon \to 0} \sup E^x[e^{-\lambda Y(S(\varepsilon))} f(X_{S(\varepsilon)})] \\ &\leqq E^x[e^{-\lambda Y(S)} f(X_S)] . \end{aligned}$$

Thus S is the optimal stopping time. From (3.17), $S(\varepsilon) \leqq T$ for each $\varepsilon > 0$ and thus, $S \leqq T$. Since S is the optimal stopping time, we also have

(3.20) $$f(X_S) = v_0(X_S) \quad \text{a.s.}$$

Since T is the first time (3.20) holds, we have $T \leqq S$, and thus, $S = T$. □

4. The optimal replacement problem. A stopping time of the form $T_\alpha = \inf\{t \geqq 0: X_t \geqq \alpha\}$, where $\alpha \in E$, is called a control limit policy. The main result of this section will be to show that a control limit policy (under suitable assumptions) is an optimal solution to Problem (2.16).

The notation of this section will be as before. Let (X, Y) be a Markov additive process; let $\zeta = \inf\{t \geqq 0: X_t = \Delta\}$; and let $\mathscr{T} = \{T \leqq \zeta: T \text{ is an } (\mathscr{M}_t) \text{ stopping time}\}$. The optimal function is given by

(4.1) $$v(x) = \sup_{T \in \mathscr{T}} E^x[e^{-\lambda Y(T)} f(X_T)] \qquad \text{for} \quad x \in E_\Delta .$$

The function given by (4.1) and the function v_0 given by (3.7) are equivalent since for $T \in \mathscr{T}_0$, $Y(T \wedge \zeta) = Y(T)$ and $f(X_{T \wedge \zeta}) = f(X_T)$.

To insure the optimality of a control limit policy, two additional assumptions are needed.

(4.2) ASSUMPTION. Let (X, Y) be a standard Markov additive process and let $f \in b\mathscr{E}_\Delta$ be the payoff function satisfying Assumption (3.3). In addition let

(a) f be concave in E_Δ;

(b) $x \to P^x\{X_t - x \geqq r, Y_t < y\}$ be nondecreasing for $r \in E$ and $t, y \geqq 0$.

REMARK. A concave function must be continuous in the interior of its domain; therefore, f is continuous in E but *not* necessarily at Δ. Condition (b) is similar to condition (3.3d) except that it involves the incremental damage instead of the absolute damage.

Theorem (3.16) actually reduces the Markov additive stopping problem to a Markov stopping problem. To see this, let $(\mathscr{K}_t)_{t\geqq 0}$ be the canonical family of σ-algebras formed from X. Define M_t^λ by

$$M_t^\lambda = E^x[e^{-\lambda Y(t)} \mid \mathscr{K}_t] \qquad \text{for} \quad t \geqq 0 . \tag{4.3}$$

By Çinlar ((1972), Proposition (2.11)), there exists a version of the conditional expectation such that $M_t^\lambda \in \mathscr{K}_t$ and M_t^λ is independent of x. Also, by Çinlar ((1972), page 104), the process $(M_t^\lambda)_{t\geqq 0}$ forms a strong multiplicative functional of the Markov process X. Therefore, $(Q_t^\lambda)_{t\geqq 0}$ defined by (3.11) can be viewed as the semigroup generated by $(M_t^\lambda)_{t\geqq 0}$. As in Blumenthal and Getoor ((1968), page 106), let $\hat{X}$ denote the Markov process formed by killing X with $(M_t^\lambda)_{t\geqq 0}$. Thus, for $x \in E$,

$$\hat{E}^x[f(\hat{X}_t)] = E^x[e^{-\lambda Y(t)} f(X_t)] + f(\Delta) E^x[1 - e^{-\lambda Y(t)}] . \tag{4.4}$$

(It should be remembered that contrary to the usual convention we let $f(\Delta) \neq 0$.)

We now prove the main result of this section.

(4.5) THEOREM. *Let $A = \{x \in E : f(x) = v(x)\}$ and assume A is not empty. Then there exists $\alpha \in E$ such that $A = [\alpha, \infty)$.*

PROOF. Let $\alpha = \inf A$. By the right-continuity of f and v, we have $\alpha \in A$. Let $u \in E$ be such that $u > \alpha$. We need to show that $u \in A$. Let $\hat{X}$ be the killed process defined above. Let $x, u \in E$ such that $x < u$. Since f is concave and $x < u$, we have for each $d > 0$

$$f(u + d) \leqq f(x + d) + f(u) - f(x) .$$

Thus,

$$\begin{aligned} \hat{E}^u[f(\hat{X}_t)] &= \hat{E}^u[f(u + (\hat{X}_t - u))] \\ &\leqq \hat{E}^u[f(x + (\hat{X}_t - u))] + f(u) - f(x) , \end{aligned}$$

combining Assumption (4.2b) and Lemma (3.4),

$$\leqq \hat{E}^x[f(x + (\hat{X}_t - x))] + f(u) - f(x) .$$

Rearranging terms and using equation (4.4),

$$E^u[e^{-\lambda Y(t)} f(X_t)] - f(u) \leqq E^x[e^{-\lambda Y(t)} f(X_t)] - f(x) + f(\Delta)\{E^u[e^{-\lambda Y(t)}] - E^x[e^{-\lambda Y(t)}]\} .$$

Since $f(\Delta) \leqq 0$ and, by Lemma (3.4), $f(\Delta)\{E^u[e^{-\lambda Y(t)}] - E^x[e^{-\lambda Y(t)}]\} \leqq 0$, thus,

$$Q_t^\lambda f(u) - f(u) \leqq Q_t^\lambda f(x) - f(x) \qquad \text{for} \quad t \geqq 0 , \quad u \geqq x . \tag{4.6}$$

Since $f(\alpha) = v(\alpha) \geqq Q_t^\lambda v(\alpha) \geqq Q_t^\lambda f(\alpha)$, equation (4.6) gives us

$$f(u) \geqq Q_t^\lambda f(u) \qquad \text{for} \quad u \geqq \alpha . \tag{4.7}$$

Define the function h by

$$h(x) = v(x) \quad \text{for} \quad x < \alpha\,,$$
$$= f(x) \quad \text{for} \quad x \geqq \alpha\,.$$

Clearly, $f \leqq h \leqq v$; so if h is λ-excessive, $h = v$ and we would be finished. In Theorem (3.14), we proved $v = g$ and in the proof of Lemma (3.13), g was shown to be right-continuous; therefore, v is right-continuous and thus, h is right-continuous. Since f is bounded by assumption, so is v and h. Using these facts,

$$\lim_{t\to 0} Q_t^{\lambda} h(x) = E^x[\lim_{t\to 0} e^{-\lambda Y(t)} h(X_t)] = h(x)\,.$$

For $x < \alpha$,

$$h(x) = v(x) \geqq E^x[e^{-\lambda Y(t)} v(X_t)]$$
$$\geqq E^x[e^{-\lambda Y(t)} h(X_t)] = Q_t^{\lambda} h(x)\,.$$

For $x \geqq \alpha$, $h(x) = f(x) \geqq Q_t^{\lambda} f(x) = Q_t^{\lambda} h(x)$ by equation (4.7). Thus, h is λ-excessive. □

(4.8) COROLLARY. *There exists an $\alpha \in E_\Delta$ such that the optimal stopping time T for Problem* (2.16) *is defined by*

$$T = \inf\{t \geqq 0 : X_t \geqq \alpha\}\,.$$

PROOF. Obvious from Theorems (3.16) and (4.5). □

Define the semi-Markov process $Z = (Z_t)_{t\geqq 0}$ using the Markov additive process (X, Y) of equations (2.11)—(2.13). The following corollary gives the optimal replacement policy for Z.

(4.9) COROLLAY. *There exists an $\alpha \in E_\Delta$ such that the optimal stopping time S for Problem* (2.15) *is defined by*

$$S = \inf\{t \geqq 0 : Z_t \geqq \alpha\}\,.$$

PROOF. By Corollary (2.18), we need only show that $S = Y_T$, but this is true since $t \longrightarrow Y_t$ is strictly increasing. □

By Corollary (4.9), the solution to the optimal replacement problem defined by (2.15) is a control limit policy. The problem becomes one of maximizing the function $\alpha \longrightarrow E^x[e^{-\lambda Y(T_\alpha)} f(X_{T_\alpha})]$ where for each $\alpha \in E$, $T_\alpha = \inf\{t \geqq 0 : X_t \geqq \alpha\}$. For E discrete, some computational considerations are given in the next section.

5. Some computational aspects. Consider a cumulative damage process Z, where Z is a standard semi-Markov process with state space $\mathbb{N}$. Let (X, Y) be the imbedded Markov additive process and, in order for replacement to make sense, assume that the lifetime ζ of X is finite almost surely. As before, let $\lambda > 0$ be a discount rate and f bounded and negative be the payoff function. Since X is a standard Markov process with a countable state space, it is a regular step process. Let $\{S_n;\, n \in \mathbb{N}\}$ be random variables with $S_0 = 0$ and $S_1, S_2, \cdots$ denoting the first, second, $\cdots$ jump times of X. Let $\hat{X}$ denote the imbedded chain, P denote the matrix of transition probabilities, and $\mu(\cdot)$ denote the sojourn rates with $a \leqq \mu(i) \leqq b$ for some $a, b > 0$ and all $i \in \mathbb{N}$; that is,

$$\begin{aligned} \hat{X}_n &= X_{S_n} && \text{for} \quad n \in \mathbb{N}\,, \\ P(i, j) &= P^i\{X_1 = j\} && \text{for} \quad i, j \in \mathbb{N}\,, \\ P^i\{S_1 > s\} &= e^{-\mu(i)s} && \text{for} \quad i \in \mathbb{N}\,. \end{aligned} \tag{5.1}$$

For notational convenience let $P(i, \Delta) = 1 - \sum_{j \in \mathbb{N}} P(i, j)$ and let the expression $\sum_{j \geq k} m(j)$ denote the infinite sum $m(k) + m(k+1) + \cdots + m(\Delta)$.

Assume that the process (X, Y) and the payoff function f satisfy Assumption (4.2). Also assume that the function $i \to P^i\{X(S_1) - i \geqq j,\, Y(S_1) < y\}$ is increasing. (Actually, (4.2) should imply this but we were unable to prove it.)

$$\begin{aligned} T_\alpha &= \inf\{t \geqq 0 : X_t \geqq \alpha\}\,, \\ h_\alpha(i) &= E^i[e^{-\lambda Y(T)} f(X_T)] \qquad \text{for} \quad i \in \mathbb{N}\,. \end{aligned} \tag{5.2}$$

It will also be convenient to define the following:

$$\begin{aligned} \tilde{P}(i, j) &= P(i, j) E^i[e^{-\lambda Y(S_1)} \mid \hat{X}_1 = j]\,, \\ \tilde{R}(i, j) &= \textstyle\sum_{n \geqq 0} \tilde{P}^n(i, j)\,. \end{aligned} \tag{5.3}$$

And thus,

$$Q^\lambda_{S_1} f(i) = \textstyle\sum_{k \geqq i+1} \tilde{P}(i, k) f(k)\,. \tag{5.4}$$

(For the Markov chain $\hat{X}$, $P(i, i) = 0$.)

(5.5) PROPOSITION. *Let $\alpha \in \mathbb{N}$ be fixed. For each $i < \alpha$, we have*

$$h_\alpha(i) = \textstyle\sum_{j < \alpha} \tilde{R}(i, j) \sum_{k \geqq \alpha} \tilde{P}(j, k) f(k)\,.$$

PROOF. Using the usual renewal theoretic argument,

$$\begin{aligned} h_\alpha(i) = &\textstyle\sum_{k \geqq \alpha} P(i, k) f(k) E^i[e^{-\lambda Y(S_1)} \mid \hat{X}_1 = k] \\ &+ \textstyle\sum_{j < \alpha} P(i, j) E^i[e^{-\lambda Y(S_1)} \mid \hat{X}_1 = j] h_\alpha(j)\,; \end{aligned}$$

thus the result follows using (5.3). □

(5.6) THEOREM. *Consider the replacement problem defined by* (2.16) *where the initial state $i \in \mathbb{N}$ is fixed. The optimal replacement time is given by equation* (5.2), *where α is the minimal value greater than or equal to i such that*

$$f(\alpha) \geqq \textstyle\sum_{k \geqq \alpha+1} \tilde{P}(\alpha, k) f(k)\,.$$

If no such α exists, then it is optimal to replace only at the failure time (*namely*, $T = \zeta$).

PROOF. The proof involves maximizing the function $\alpha \to h_\alpha(i)$. Consider the difference $\Delta h_\alpha(i) = h_{\alpha+1}(i) - h_\alpha(i)$. We first find an α such that the difference is nonpositive and then show that if the difference is nonpositive for some α, it will remain nonpositive for all values larger than α. Because the Markov chain $\hat{X}$ is increasing, $P(i, j) = 0$ for $i > j$; furthermore, $P(i, i) = 0$ since $\hat{X}$ was imbedded in a standard Markov process. Therefore, we have

$$\tilde{R}(i, i) = 1 , \tag{5.7}$$

$$\sum_{j<\alpha} \tilde{R}(i, j)\tilde{P}(j, \alpha) = \tilde{R}(i, \alpha) - I(i, \alpha) . \tag{5.8}$$

Using (5.2), (5.5) and (5.7), it is easy to see that

$$\Delta h_i(i) = \sum_{k\geqq i} \tilde{P}(i, k)f(k) - f(i) .$$

For $\alpha > i$, after simplifying and using (5.8), we obtain

$$\Delta h_\alpha(i) = \tilde{R}(i, \alpha) \sum_{k\geqq \alpha+1} \tilde{P}(\alpha, k)f(k) - f(\alpha)\tilde{R}(i, \alpha) .$$

Solving the expression $\Delta h_\alpha(i) \leqq 0$, the equation of the theorem is obtained. For the local maximum to be the global maximum, it is sufficient to show that $\alpha \to \Delta h_\alpha(i)$ is decreasing. Using (5.4) it is seen that we need show that $\alpha \to Q^\lambda_{S_1} f(\alpha) - f(\alpha)$ is decreasing. Using the additional assumption that $i \to P^i\{X(S_1) - i \geqq j, Y(S_1) < y\}$ is increasing the desired result follows by Lemma (3.4) and using the derivation of (4.6).

Computationally, the difficult part of Theorem (5.6) is calculating the expression $E^i[e^{-\lambda Y(S_1)} \mid \hat{X}_1 = j]$. To put this in a manageable form the decomposition of Y as given in Proposition (2.9) is used. The notation of Çinlar (1974) is followed.

From (2.10) we write (remembering that $Y^p = 0$)

$$Y = A^c + Y^q + Y^d .$$

For each $i \in \mathbb{N}$, there is a positive real number a_i such that

$$A_t^c = a_i t \qquad \text{on} \quad \{X_0 = i, S_1 > t\} .$$

For each $i, j \in \mathbb{N}$, there exists a distribution function $F(i, j, \cdot)$ with support in $\mathbb{R}_+$ such that

$$P^i\{Y_t^q \leqq y \mid \hat{X}_1 = j, S_1 = t\} = F(i, j, y) .$$

And finally for each $i \in \mathbb{N}$, there exists a Lévy measure ν_i such that

$$E^i[e^{-\lambda Y_t^d} \mid \{S_1 > t\}] = \exp\{t \textstyle\int_{(0,\infty)} (1 - e^{-\lambda y})\nu_i(dy)\} .$$

For notational convenience, let

$$\lambda N_i^\lambda = \lambda a_i + \int_{(0,\infty)} (1 - e^{-\lambda y})\nu_i(dy)$$
$$F^\lambda(i, j) = \int e^{-\lambda y} F(i, j, dy) .$$

Then

$$E^i[e^{-\lambda Y(S_1)} \mid \hat{X}_1 = j] = E^i[e^{-\lambda(A^c(S_1) Y^d(S_1))}] + E^i[e^{-\lambda Y^q(S_1)} \mid X_1 = j]$$
$$= \frac{\mu_i}{\lambda N_i^\lambda + \mu_i} + F^\lambda(i, j) .$$

Thus

$$\tilde{P}(i, j) = P(i, j)[F^\lambda(i, j) + \mu_i/(\lambda N_i^\lambda + \mu_i)] . \tag{5.9}$$

Equation (5.9) can now be used in Theorem (5.6) to obtain a closed form expression for the optimal replacement level using known distribution functions.

Acknowledgment. I would like to thank Professor Erhan Çinlar for his continual guidance and advice and to thank the referees for their many excellent and helpful suggestions.

REFERENCES

[1] BLUMENTHAL, R. M. and GETOOR, R. K. (1968). *Markov Processes and Potential Theory.* Academic Press, New York.

[2] DELLACHERIE, C. (1972). *Capacités et Processus Stochastiques.* Springer-Verlag, Berlin.

[3] ÇINLAR, E. (1972). Markov additive processes II. *Z. Wahrscheinlichkeitsthorie und Verw. Gebiete* **24** 95–121.

[4] ÇINLAR, E. (1974). Markov additive processes and semi-regeneration. The Center for Mathematical Studies in Economics and Management Science, Discussion paper no. 118, Northwestern Univ.

[5] FAKEEV, A. G. (1970). Optimal stopping rules for stochastic processes with continuous parameter. *Theor. Probability Appl.* **15** 324–331.

[6] FAKEEV, A. G. (1971). Optimal stopping of a Markov process. *Theor. Probability Appl.* **16** 694–696.

[7] FELDMAN, R. M. (1976a). Optimal replacement with semi-Markov shock models. *J. Appl. Probability* **13** 108–117.

[8] FELDMAN, R. M. (1976b). Optimal replacement with semi-Markov shock models using discounted costs. *Math. Operations Res.* To appear.

[9] JACOD, J. (1974). Systèmes régénératifs et processus semi-Markoviens. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31** 1–23.

[10] MAISONNEUVE, B. (1974). Systèmes régénératifs. *Asterisque* (Soc. Math. France) **15** 1–123.

[11] ROYDEN, H. L. (1968). *Real Analysis.* Collier-Macmillan, London.

[12] SERFOZO, R. F. (1977). Monotone optimal policies for Markov decision processes. In *Stochastic Systems: Modeling Identification, and Optimization* (R. Wets, ed.). North-Holland, Amsterdam (in press).

[13] TAYLOR, H. M. (1968). Optimal stopping in a Markov process. *Ann. Math. Statist.* **39** 1333–1344.

DEPARTMENT OF INDUSTRIAL ENGINEERING
TEXAS A & M UNIVERSITY
COLLEGE STATION, TEXAS 77843