The Annals of Statistics
1975, Vol. 3, No. 2, 547-553

MARKOV DECISION PROCESSES WITH A NEW OPTIMALITY CRITERION: CONTINUOUS TIME

BY STRATTON C. JAQUETTE

Cornell University

Standard finite state and action continuous time Markov decision processes with discounting are studied using a new optimality criterion called moment optimality. A policy is moment optimal if it lexicographically maximizes the sequence of signed moments of total discounted return with a positive (negative) sign if the moment is odd (even). It is shown constructively that a stationary policy is moment optimal among the class of piecewise constant policies by examining the negative of the Laplace transform of the total return random variable and its Taylor series expansion.

1. Introduction. This paper is concerned with finite state and action continuous time Markov decision processes where future returns are discounted. We study a new optimality criterion, moment optimality, for continuous time processes and follow the development for the discrete time case given in Jaquette (1973). This paper incorporates some of the results of the author's dissertation (1971).

The Markov decision processes as well as most of the definitions and notation needed to develop the results of this paper are given in Section 2. Section 3 contains the result that there is a stationary policy which is moment optimal among the class of stationary policies. The method of proof is constructive. In Section 4 we extend these results to show that stationary policies are moment optimal in the wider class of piecewise constant policies.

2. Preliminaries and notations. We consider a standard Markov decision process. We assume that the stochastic process has a finite state space denoted by S, $S = \{1, 2, \cdots, s\}$. The process starts at time $t = 0$ and can jump from state to state at any point in time. At each instant of time an action is selected and applied to the Markov process. We assume that there are perhaps distinct finite action sets A_i available only when the process is in state i but at any instant in time. We define the set F by $F = \times_{i=1}^{s} A_i$. We call the elements of F action vectors and let $\mathbf{f}$ denote an element of F. The ith component of $\mathbf{f}$, denoted $f(i)$, is the action taken if the process is in state i.

A policy is a mapping from the time axis into the set F. If π is a policy, then $\pi(t)$ is the action vector used at time t. We only admit measurable policies in the discussions here; the policy π is admissible if the set $\{t : \pi(t) = \mathbf{f}\} = \pi^{-1}(\mathbf{f})$ is

Received October 1971; revised April 1974.

AMS 1970 *subject classifications.* Primary 90B99, 60J25, 93E20; Secondary 90C40, 90B99.

Key words and phrases. Dynamic programming, Markov decision processes, optimality criterion, moments of return.

a Lebesgue measurable set of the nonnegative real line for each $\mathbf{f} \in F$. We let f^∞ denote the stationary policy such that $f^\infty(t) = \mathbf{f}$ for all t and let $X_\pi(t)$ denote the random state of the process at time t when the policy π is used.

The transition probabilities are also determined by the policy. If $\mathbf{f} \in F$ is the action vector used at some point in time, then Q_f will represent the infinitesimal generator of the Markov process at that point in time. We let $Q_f = \{q(j \mid i, f(i))\}$ and assume that $0 \leqq q(j \mid i, a) < \infty$ for $j \neq i$ and that $\sum_{j=1}^{s} q(j \mid i, a) = 0$ for any $a \in A_i$.

Infinitesimal generators have been studied extensively; we use results discussed in Miller (1968) and Dynkin (1961). We know that for any measurable policy there exists a unique absolutely continuous Markov transition function satisfying $P(s, t) = I + Q_{\pi(s)}(t - s) + o(t - s)$ for almost all $s(t > s)$. $P(s, t)$ is the matrix of transition probabilities, $P(s, t) = \{p_{ij}\}$, where p_{ij} is the probability that the process will be in state j at time t given that the process was in state i at time s and the policy π is used in the interval (s, t). We also know that $(d/dt)P(s, t) = P(s, t)Q_{\pi(t)}$ for almost all t.

The action applied to the process at any point in time determines the rate at which returns are earned and the transition probability rates at that point in time. We denote by $r(i, f(i))$ the rate of return earned when the process is in state i and action $f(i)$ is applied. Thus $\mathbf{r}_f$ is the vector with ith component $r(i, f(i))$. We assume discounting so that, if $\alpha(\alpha > 0)$ is the discounting rate, a unit of return earned at time t has present worth $e^{-\alpha t}$. The total discounted return random variable, $\mathbf{R}(\pi)$, is defined by:

$$\text{(2.1)} \qquad \mathbf{R}(\pi) = \textstyle\int_0^\infty \mathbf{r}_{\pi(t)} \circ \mathbf{X}_\pi(t) e^{-\alpha t}\, dt\,,$$

where $\mathbf{X}_\pi(t)$ is the vector whose ith component is 1 if $X_\pi(t) = i$ and 0 otherwise. The composition operator $\circ$ indicates componentwise multiplication of vectors of the same dimension, i.e. $\mathbf{u} \circ \mathbf{v}$ has ith component $u_i v_i$.

We can characterize the total discounted return random vector $\mathbf{R}(\pi)$ given in (2.1) by its moments. Let $\mathbf{M}_n(\pi)$ be the nth moment of $\mathbf{R}(\pi)$ with $\mathbf{M}_n(\pi) \equiv E[\mathbf{R}(\pi)^n]$ for $n = 1, 2, \cdots$ and $\mathbf{M}_0(\pi) \equiv \mathbf{1}$ where $\mathbf{R}(\pi)^n = \mathbf{R}(\pi) \circ \mathbf{R}(\pi) \circ \cdots \circ \mathbf{R}(\pi)$. We also consider vectors $\mathbf{N}_n(\pi)$: $\mathbf{N}_n(\pi) \equiv (-1)^{n+1}\mathbf{M}_n(\pi)$ $(n = 0, 1, 2, \cdots)$ and use the collection of vectors $\mathbf{N}(\pi)$:

$$\mathbf{N}(\pi) \equiv (\mathbf{N}_0(\pi), \mathbf{N}_1(\pi), \cdots, \mathbf{N}_n(\pi), \cdots)\,.$$

We also use the negative of the Laplace transform of the total discounted return random vector, $\mathbf{R}$, which we define as

$$\text{(2.2)} \qquad \mathbf{U}_\pi(\lambda) \equiv -E[\exp(-\lambda \mathbf{R}(\pi))]\,.$$

From our definitions it is clear that $\mathbf{R}(\pi)$ is bounded for all policies and that therefore its Laplace transform exists everywhere. Also since the Laplace transform is analytic on the interior of its region of convergence, being here the real line, it follows directly that $\mathbf{U}_\pi(\lambda)$ exists everywhere with the Taylor series expansion given by $\mathbf{U}_\pi(\lambda) = \sum_{n=0}^{\infty} \mathbf{N}_n(\pi)\lambda^n/n!$. Thus we can use $\mathbf{U}_\pi(\lambda)$ or $\mathbf{N}(\pi)$ interchangeably.

We use the complete ordering of vectors, lexicographic ordering, and a slight generalization of its usual application. Suppose $\mathbf{N}$ and $\mathbf{M}$ are vectors whose components are themselves vectors. In this case we write $\mathbf{N} \succ \mathbf{M}$ if there is an integer n such that $\mathbf{N}_i = \mathbf{M}_i$ for $i < n$ and $\mathbf{N}_n > \mathbf{M}_n$. We define the relationships $\succeq$, $\prec$, and $\preceq$ in the obvious fashion.

It will be useful to use $\mathbf{N}$ to represent the collection of vector coefficients $\mathbf{N}_n$ in the Taylor series expansion of $\mathbf{U}(\lambda) = \sum_{n=0}^{\infty} \mathbf{N}_n(\lambda^n/n!)$. It will also be convenient to use the notation $\mathbf{N} =_m \mathbf{M}$ to mean that $\mathbf{N}_n = \mathbf{M}_n$ for $n \leqq m$. By $\mathbf{N} \succ_m \mathbf{M}$ we mean $\mathbf{N} \neq_m \mathbf{M}$ and $\mathbf{N} \succ \mathbf{M}$.

We say that a policy π^* is *moment optimal* if $\mathbf{N}(\pi^*) \succeq \mathbf{N}(\pi)$ for all policies π. It is also convenient to note that $\mathbf{N}(\pi^*) \succeq \mathbf{N}(\pi)$ if and only if $\mathbf{U}_{\pi^*}(\lambda) \geqq \mathbf{U}_\pi(\lambda)$ for all λ in $[0, \lambda_0]$ for some $\lambda_0 > 0$. This is easy to verify as done in Jaquette (1973), Lemma 3.6.

The implications of moment optimality are discussed in detail in Jaquette (1971) and (1973). If a policy is moment optimal, it yields greatest expected return, minimum second moment or variance among all policies yielding the greatest expected return, etc. A moment optimal policy is also the optimal policy which maximizes expected utility of return using an exponential utility function with suitably low aversion to risk.

We now define several operators and spaces that will be used in our subsequent development. Define $\mathscr{L}$ to be the space of analytic functions of the form $\sum_{n=0}^{\infty} \mathbf{N}_n(\lambda^n/n!)$ where the $\mathbf{N}_n$ are s dimensional vectors, $\mathbf{N}_0 = -\mathbf{1}$, and the sum converges for all λ. Define the mapping $L(f, t)$, which takes $\mathscr{L}$ into itself, as

$$(2.3)\qquad (L(f, t)\mathbf{U})(\lambda) \equiv E[\exp(-\lambda \textstyle\int_0^t \mathbf{r}_f \circ \mathbf{X}_{f\infty}(\tau)e^{-\alpha\tau}\, d\tau)] \circ P(t, f)\mathbf{U}(e^{-\alpha t}\lambda)\,,$$

where $\mathbf{U}(\lambda)$ is an element of $\mathscr{L}$ and $P(t, f)$ is the probability transition matrix for time t using the policy f^∞. It is not hard to verify that $L(f, t)$ maps $\mathscr{L}$ into itself.

The mapping $L(f, t)$ is monotone, which is simple to verify using the definition (2.3). Suppose $\mathbf{U}_1, \mathbf{U}_2 \in \mathscr{L}^s$ and $\mathbf{U}_1(\lambda) \leqq \mathbf{U}_2(\lambda)$ for $\lambda \in [0, \lambda_0]$, then by examining $(L(f, t)(\mathbf{U}_1 - \mathbf{U}_2))(\lambda)$ we see that $L(f, t)\mathbf{U}_1(\lambda) \leqq L(f, t)\mathbf{U}_2(\lambda)$ for $\lambda \in [0, \lambda_0]$.

It is laborious, but not difficult to verify that

$$\mathbf{U}_{f^t\pi}(\lambda) = (L(f, t)\mathbf{U}_\pi)(\lambda)\,,$$

where the policy $f^t\pi$ is the policy that uses $\mathbf{f}$ until time t and then switches to π. This requires expanding $\mathbf{U}_{f^t\pi}(\lambda)$ using (2.1) and (2.2), rewriting this expectation by conditioning on the sample path from time 0 to t, judiciously removing the conditioning where it is not relevant, reducing the conditioning to $\mathbf{X}(t)$ in other places, and observing that the result corresponds to the definition of $L(f, t)\mathbf{U}_\pi(\lambda)$. This allows us to conclude that $L(f, t_1)L(f, t_2) = L(f, t_1 + t_2)$.

We also define a criterion operator θ_g which acts on elements of $\mathscr{L}$. For any $\mathbf{g}$ and $\mathbf{U}(\lambda)$ in $\mathscr{L}$ define

$$\theta_g \mathbf{U}(\lambda) \equiv \frac{d}{dt} L(g, t)\mathbf{U}(\lambda)\Big|_{t=0} .$$

This can be evaluated by laborious formal differentiation of (2.3) and evaluating at $t = 0$. The validity of this procedure is a consequence of an application of the dominated convergence theorem; cf. Loève (1960) page 126. The result is

$$\theta_g \mathbf{U}(\lambda) = -\lambda \mathbf{r}(g) \circ \mathbf{U}(\lambda) + Q_g \mathbf{U}(\lambda) - \alpha\lambda \frac{d}{d\lambda} \mathbf{U}(\lambda) . \tag{2.4}$$

Note that by definition $\theta_f \mathbf{U}_{f^\infty}(\lambda) = \mathbf{0}$ since $L(f, t)\mathbf{U}_{f^\infty}(\lambda) = \mathbf{U}_{f^\infty}(\lambda)$.

3. Constructing the best stationary policy. We first restrict attention to stationary policies and show that there is a stationary policy that is moment optimal within the class of stationary policies. Our proof is by construction. Since stationary policies will turn out to be optimal in a more general setting, we allow the following proof to stand for an algorithm to construct a moment optimal policy.

In this section we will use a criterion called (m) moment optimality. A policy π^* is called *(m) moment optimal* if $\mathbf{N}(\pi^*) \succeq_m \mathbf{N}(\pi)$ for all policies π. It should be evident that a policy is moment optimal if and only if it is (m) moment optimal for all m.

Define the sets of action vectors, F^m, as follows:

$$F^m \equiv \{\mathbf{f} : f^\infty \text{ is } (m) \text{ moment optimal among stationary policies}\} .$$

The set of action vectors, $\mathbf{f}$, such that f^∞ is moment optimal is simply $\lim_{m\to\infty} F^m$ which we denote by F^∞. If F^m is nonempty for each m, then F^∞ is nonempty and there are moment optimal policies.

In the following lemma and corollary we characterize when the moments of return for two policies coincide for the first m moments. This also allows us to characterize F^m, the set of action vectors generating stationary (m) moment optimal policies. In this development we let $\theta_f \mathbf{U}$ represent the vector collection of Taylor series coefficients for $\theta_f \mathbf{U}(\lambda)$.

LEMMA 3.1. *Let* $\mathbf{f}$ *be fixed and* $\mathbf{U}(\lambda)$ *be an element of* $\mathscr{L}$. *Then* $\mathbf{N}(f^\infty) =_m \mathbf{N}$, $\theta_f \mathbf{U} =_m \mathbf{0}$, *and* $\mathbf{N}_n = (Q_f - n\alpha I)^{-1}(n\mathbf{r}_f \circ \mathbf{N}_{n-1})$ *for* $1 \leqq n \leqq m$ *are equivalent statements.*

PROOF. Examining the expansion for $\mathbf{U}(\lambda)$ in (2.4) and observing that $(Q_f - n\alpha I)$ is nonnegative off the diagonal and nonpositive on the diagonal insures that $(Q_f - n\alpha I)^{-1}$ exists and therefore that $\theta_f \mathbf{U} =_m \mathbf{0}$ and $\mathbf{N}_n = (Q_f - n\alpha I)^{-1}(n\mathbf{r}_f \circ \mathbf{N}_{n-1})$ $(1 \leqq n \leqq m)$ are equivalent. Since $\theta_f \mathbf{U}_{f^\infty} = \mathbf{0}$, $\mathbf{N}_n(f^\infty) = (Q_f - n\alpha I)^{-1}(n\mathbf{r}_f \circ \mathbf{N}_{n-1}(f^\infty))$ for all n. This shows the equivalence of the first and last statements in the lemma.

COROLLARY 3.2. *If* $\mathbf{f}$ *is an element of* F^m, *then* $F^m = \times_{i=1}^{s} A_i^m$ *where* $A_i^m = \{a \in A_i^{m-1}: \theta_g \mathbf{U}_{f^\infty} =_m \mathbf{0}$ *for any* $\mathbf{g}$ *such that* $\mathbf{g} = \mathbf{f}$ *except* $g(i) = a\}$.

PROOF. Lemma 3.1 says that $\theta_g \mathbf{U}_{f^\infty} =_m \mathbf{0}$ characterizes all $\mathbf{g}$ for which the m moments of g^∞ match those of f^∞. This then characterizes all $\mathbf{g}$ in F^m given that f is in F^m. Restricting F^m to be a subset of F^{m-1}, this condition and (2.5) gives

$$(3.1) \qquad (Q_g - m\alpha I)\mathbf{N}_m(f^\infty) - m\mathbf{r}_g \circ \mathbf{N}_{m-1}(f^\infty) = \mathbf{0}\,.$$

In this form it is clear that the selection of $g(i)$ completely determines the ith component of the left side of (3.1), and thus that F^m can be chosen as the Cartesian product indicated.

Policy improvement requires more arguments. Our policy improvement procedure is based on the fact that the Cartesian product nature of F^{m-1} allows componentwise selection of f to maximize $\mathbf{N}_m(f^\infty)$. First we need the following lemma.

LEMMA 3.3. *Let* $\mathbf{f}$ *be fixed and* $\mathbf{U}(\lambda)$ *an element of* $\mathscr{L}$ *with Taylor expansion coefficients* $\mathbf{N}$. *If* $\theta_f \mathbf{U} \succ \mathbf{0}$, *then* $\mathbf{N}(f^\infty) \succ \mathbf{N}$.

We remark that this lemma is true with the other lexicographic orderings replacing $\succ$ in hypothesis and conclusion.

PROOF. To demonstrate this result we must go through an intermediate examination of $L(f, t)\mathbf{U}(\lambda)$. Suppose $(d/dt)L(f, t)\mathbf{U}(\lambda) > \mathbf{0}$ for all t and λ such that $0 < t \leqq t_0$, $0 < \lambda \leqq \lambda_0$ for some λ_0 and t_0. Then $L(f, t)\mathbf{U}(\lambda) > \mathbf{U}(\lambda)$ for all t in $(0, t_0]$ and all λ in $(0, \lambda_0]$. By monotonicity and repeated use of $L(f, t)$, this would mean that $L(f, t)\mathbf{U}(\lambda) > \mathbf{U}(\lambda)$ for all t and all λ in $(0, \lambda_0]$. $L(f, t)\mathbf{U}(\lambda)$ converges to $\mathbf{U}_{f^\infty}(\lambda)$ as $t \to \infty$, and we can conclude that $\mathbf{U}_{f^\infty}(\lambda) > \mathbf{U}(\lambda)$ for all λ in $(0, \lambda_0]$. This suffices to show that $\mathbf{N}(f^\infty) \succ \mathbf{N}$. Also observe that $(d/dt)L(f, t) = L(f, t)\theta_f$. Now by assumption $\theta_f \mathbf{U} \succ \mathbf{0}$, hence $\theta_f \mathbf{U}(\lambda) > \mathbf{0}$ for all λ in some interval close to zero, say in $(0, \lambda_0]$ with $\lambda_0 > 0$. This implies that $(d/dt)L(f, t)\mathbf{U}(\lambda) > \mathbf{0}$ for t and λ close to zero and suffices for the proof. For further technical details see Jaquette (1971), Lemma 7.5 and Jaquette (1973).

LEMMA 3.4. *Choose any* $\mathbf{f}$ *in* F^{m-1}. *Either*

(a) $\mathbf{f} \in F^m$ *or*

(b) *there exists a* $\mathbf{g} \in F^{m-1}$ *such that* $\mathbf{N}(g^\infty) \succ_m \mathbf{N}(f^\infty)$.

PROOF. The proof is familiar and follows closely that of Lemma 4.3 in Jaquette (1973). For all $\mathbf{g}$ in F^{m-1} $\mathbf{N}(g^\infty) =_{m-1} \mathbf{N}(f^\infty)$ and hence by Lemma 3.1 $\theta_g \mathbf{U}_{f^\infty} =_{m-1} \mathbf{0}$. If there exist $\mathbf{g}$ in F^{m-1} such that $\theta_g \mathbf{U}_{f^\infty} \neq_m \mathbf{0}$, then the methods of Corollary 3.2 allow us to choose a $\mathbf{g}$ which maximizes $[\theta_g \mathbf{U}_{f^\infty}]_m$ componentwise. For this $\mathbf{g}$ either $\theta_g \mathbf{U}_{f^\infty} =_m \mathbf{0}$ or $\theta_g \mathbf{U}_{f^\infty} \succ_m \mathbf{0}$. In the former case f is in F^m. This follows since for all $\mathbf{g}$ in F^{m-1} either $\theta_g \mathbf{U}_{f^\infty} \prec_m \mathbf{0}$ and hence $\mathbf{U}_{f^\infty} \succ_m \mathbf{U}_{g^\infty}$ by a version of Lemma 3.3 or $\theta_g \mathbf{U}_{f^\infty} =_m \mathbf{0}$ and hence $\mathbf{U}_{f^\infty} =_m \mathbf{U}_{g^\infty}$ by Lemma 3.1. In the latter case Lemma 3.3 insures (b) of the lemma.

We now state the major result of this section.

THEOREM 1. *There is a stationary policy which is moment optimal within the class of stationary policies. An optimal policy can be constructed in a finite number of steps.*

PROOF. Start with $F^0 = \bigtimes_{i=1}^{s} A_i^0$, where $A_i^0 = A_i$. In general Lemma 3.4 indicates a construction to obtain an element of F^m given F^{m-1}. This construction terminates in a finite number of steps with an element of F^m since F is finite. Corollary 3.2 indicates a construction of F^m given this one element which again is finite. Thus all stationary (m) moment optimal policies can be constructed. Theorem 3 in Jaquette (1973) insures that there is a finite number n^* such that (n^*) moment optimality is equivalent to moment optimality, this since the number of stationary policies is finite. This completes our proof.

Operationally we need a way of determining when to stop, i.e. of determining what n^* is. A stopping rule is quite clear: stop at the smallest m, equal to n^*, where $\mathbf{R}(f^\infty) =_{\mathscr{D}} \mathbf{R}(g^\infty)$ whenever $\mathbf{f}, \mathbf{g} \in F^m$. This will of course be true if F^m reduces to a single element.

4. Piecewise constant policies do no better. In Section 3 we showed that within the class of stationary policies there exists one which is moment optimal. We now consider a larger class of policies and show that piecewise constant policies cannot improve upon the stationary moment optimal policy found in Section 3. We do this in the following lemmas.

LEMMA 4.1. *Suppose a policy π^* and numbers $\lambda_0 > 0$ and $t_0 > 0$ exist such that*

$$(4.1) \qquad \mathbf{U}_{\pi^*}(\lambda) \geqq \mathbf{U}_{g^t\pi^*}(\lambda) \qquad \textit{for all} \quad \mathbf{g} \in F\,, \quad \lambda \in [0, \lambda_0]\,, \quad \textit{and} \quad t \in [0, t_0]\,.$$

Then $\mathbf{U}_{\pi^*}(\lambda) \geqq \mathbf{U}_\pi(\lambda)$ for all $\lambda \in [0, \lambda_0]$ aud all piecewise constant policies π.

PROOF. Write $\mathbf{U}_{g^t\pi^*}(\lambda)$ as $L(g, t)\mathbf{U}_{\pi^*}(\lambda)$. By repeated application of $L(g, t)$ on (4.1) for various $t \leqq t_0$ it is clear that one can obtain $\mathbf{U}_{\pi^*}(\lambda) \geqq L(g, t)\mathbf{U}_{\pi^*}(\lambda)$ for any desired t. Thus we can choose $t_0 = +\infty$ and dispense with this restriction.

Now choose any piecewise constant policy, π, and any interval of time $[0, T]$. There is then a finite sequence of times and action vectors, $\{t_i\}$ $(0 < t_1 < \cdots < t_m = T)$ and $\{\mathbf{f}_i\}$ $(i = 1, \cdots, m)$, such that $\pi(t) = \mathbf{f}_i$ whenever $t \in (t_{i-1}, t_i)$. It is immaterial whether $\pi(t_i)$ equals $\mathbf{f}_i$ or $\mathbf{f}_{i+1}$. Let $\pi_T = f_1^{t_1} f_2^{t_2-t_1} \cdots f^{T-t_{m-1}}\pi^*$. Since

$$(L(f_1, t_1)(L(f_2, t_2 - t_1) \cdots (L(f_m, T - t_{m-1})\mathbf{U}_{\pi^*}) \cdots))(\lambda) = \mathbf{U}_{\pi_T}(\lambda)\,,$$

m is finite, and $L(f, t)$ is monotone for any t and $\mathbf{f}$, we can apply the operators $L(f_i, t_i - t_{i-1})$ in turn and preserve the inequality of (4.1) for all λ in $[0, \lambda]$. It then follows that

$$\mathbf{U}_{\pi^*}(\lambda) \geqq \mathbf{U}_{\pi_T}(\lambda) \qquad \text{for all} \quad \lambda \in [0, \lambda_0]\,, \quad \text{any} \quad \pi\,, \quad \text{and} \quad T\,.$$

We now let $T \to \infty$. It can easily be shown that $\mathbf{U}_{\pi_T}(\lambda) \to \mathbf{U}_\pi(\lambda)$ as $T \to \infty$ for all λ, and we may conclude that $\mathbf{U}_{\pi^*}(\lambda) \geqq \mathbf{U}_\pi(\lambda)$ for $\lambda \in [0, \lambda_0]$ and any τ. This is just the required result.

LEMMA 4.2. *A stationary policy is moment optimal within the class of piecewise constant policies.*

PROOF. Let the stationary policy found in Section 3, Theorem 1 be f^∞. It satisfies $\theta_g \mathbf{U}_{f^\infty} \preceq \mathbf{0}$ for all $\mathbf{g}$. By essentially the same arguments given in Lemma 3.3, this implies that $L(g, t)\mathbf{U}_{f^\infty}(\lambda) \leqq \mathbf{U}_{f^\infty}(\lambda)$ for all $\mathbf{g}$ and all positive λ and t small enough. This is simply (4.1), which implies that no piecewise constant policy can improve on f^∞. Note that the fact that such a $\lambda_0 > 0$ and $t_0 > 0$ follow from the finiteness of F.

Acknowledgments. The author wishes to acknowledge the assistance of his advisor, Professor Donald L. Iglehart, while at Stanford University, and the helpful suggestions of Professor Eric V. Denardo indicating a more compact presentation of these results.

REFERENCES

[1] DYNKIN, E. B. (1961). *Theory of Markov Processes.* Prentice Hall, Englewood Cliffs, New Jersey. (Translated from Russian).

[2] JAQUETTE, S. C. (1971). Markov decision processes with a new optimality criterion. Technical Report No. 15, Department of Operations Research, Stanford Univ.

[3] JAQUETTE, S. C. (1973). Markov decision processes with a new optimality criterion: discrete time. *Ann. Statist.* **1** 496–505.

[4] LOÈVE, M. (1960). *Probability Theory* (2nd ed.). Van Nostrand, New York.

[5] MILLER, B. L. (1968). Finite state continuous-time Markov decision processes with a finite planning horizon. *SIAM J. Control* **6** 266–280.

DEPARTMENT OF OPERATIONS RESEARCH
UPSON HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853