# AN ABSTRACT OF THE THESIS OF

MELVIN LEROY OTT for the degree DOCTOR OF PHILOSOPHY
(Name) (Degree)

in STATISTICS presented on August 9, 1974
(Major Department) (Date)

Title: OPTIMAL POLICIES IN CONTINUOUS MARKOV DECISION CHAINS

Redacted for privacy

Abstract approved: 
Mark R. Lembersky

For continuous time, finite state and action, Markov decision chains, optimal policies are studied; (i) a procedure for transforming the terminal reward vector is given and it is established that this transformation does not alter optimal policies, (ii) decision chains with absorbing states are studied and the results obtained are applied to an environmental control problem to find an optimal policy, (iii) under state accessibility and recurrence hypotheses the set of preferred decision rules is obtained, and (iv) conditions for stationary and initially stationary policies to attain maximal expected rewards on various recurrent states are given in terms of the limit vector V.

# Optimal Policies in Continuous Markov Decision Chains

by

Melvin Leroy Ott

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1975

APPROVED:

Redacted for privacy

Assistant Professor of Statistics
in charge of major

Redacted for privacy

Chairman of Department of Statistics

Redacted for privacy

Dean of Graduate School

Date thesis is presented August 9, 1974

Typed by Clover Redfern for Melvin Leroy Ott

# ACKNOWLEDGMENT

The author would like to express his appreciation to Dr. Mark Lembersky who suggested and directed the research in this paper. Without his encouragement and valuable suggestions this thesis would not have been possible.

Thanks must also be given to Diantha (my wife) for her support and understanding during the writing of this thesis.

# TABLE OF CONTENTS

# OPTIMAL POLICIES IN CONTINUOUS MARKOV DECISION CHAINS

## INTRODUCTION AND SUMMARY

This paper is concerned with continuous time, finite state and action, Markov decision chains.

Chapter I describes the process, reviews some of the current literature, and presents some preliminaries. A transformation is introduced in the final section of Chapter I and using this transformation it will be established that the terminal reward vector may be transformed without altering optimal policies.

In Chapter II decision chains with absorbing states are first discussed and the results obtained are applied to an environmental control problem.

The next section deals with $\epsilon$-optimal policies, and under a recurrence hypothesis, a result is given for finding the set of decision rules that may be used as the stationary segments in initially stationary $\epsilon$-optimal policies. The value of $\epsilon$-optimal policies is further shown by an example in which there does not exist an initially stationary optimal policy.

Stationary and initially stationary optimal policies are studied in Section 4 of Chapter II. The results obtained provide necessary and sufficient conditions for a stationary policy to give maximal expected rewards on various recurrent states.

Referencing of results stated in a different chapter will be done by giving the chapter number following the designation of the result, i.e., Lemma 2, I, indicates Lemma 2 of Chapter I. For results in the same chapter, the chapter reference will be omitted.

# I. PRELIMINARY RESULTS

## 0. Introduction

The system will be introduced in Section 1 while Sections 2, 3 and 4 will be devoted to giving additional structure plus some of the known results in the literature. The formulation used here follows the same notation as that of Lembersky [12].

The final section deals with the vector of terminal rewards. Using a transformation, it is possible to replace the terminal reward vector in $R^N$ with any other vector in $R^N$ and not alter the optimal policies. This result will be of special value for converting the terminal reward to zero.

## 1. The System

Consider a system that is always in one of $N$ states, labeled $S = \{1, 2, \ldots, N\}$. When the system is in state $i$, as action $a$ is selected from the finite set $A_i$, an $A_i$ being defined for each $i \in S$. For each action $a \in A_i$ there is a set of transition rates $\{q(j|i, a), j \in S\}$ and a reward rate $r(i, a)$. The transition rates satisfy

$$\sum_{j=1}^{N} q(j|i,a) = 0 \quad \text{and} \quad q(j|i,a) \geq 0 \quad \text{for} \quad j \neq i.$$

Let $F = \overset{N}{\underset{i=1}{\times}} A_i$ and say $f \in F$ is a <u>decision rule</u>. Then for each $f \in F$, let $Q(f)$ denote the $N \times N$ Markov infinitesimal generator matrix whose ijth element is $q(j|i, f(i))$ and let $r(f)$ be the $N \times 1$ column vector whose ith component is $r(i, f(i))$.

A <u>policy</u> $\pi: [0, \infty) \to F$ is any measurable function which specifies for each $t \geq 0$ a decision rule in $F$. The policy $\pi$ being measurable means that for every $f \in F$, $\{t \geq 0: \pi(t) = f\}$ is a Lebesgue measurable subset of $[0, \infty)$. Note that $\pi$ describes the actions to be selected for every possible combination of times and states of the system. A policy $\pi$ is defined on the entire interval $[0, \infty)$, but in a decision process of duration $t$ using $\pi$, only the decision rules $\pi(s)$, $0 \leq s \leq t$ are used.

In the current discussion take $[0, \infty)$ as the time index and reverse time so that, for a process of duration $t \geq 0$, the process begins at time $t$ and the time index decreases to zero.

From Miller [14], for each policy $\pi$, the set $\{Q(\pi(t)), t \geq 0\}$ determines a continuous time Markov process, with piecewise constant sample paths, and with transition function $P(t, s; \pi)$, for all $t \geq s \geq 0$. Also for each $t \geq 0$, $P(t, s; \pi)$ is the unique, absolutely continuous in $s$ matrix function satisfying

$$\frac{-\partial}{\partial s} P(t, s; \pi) = P(t, s; \pi) Q(\pi(s)) \tag{1}$$

for almost all $0 \leq s \leq t$, with initial condition $P(t,t;\pi) = I$. The $ij$th element of $P(t,s;\pi)$ is the probability that the system is in state $j$ at time $s$ ($\leq t$), given that it was in state $i$ at time $t$ and that the policy $\pi$ is being used.

For $P(t,0;\pi)$, an abbreviated notation will be used, namely $P(t;\pi)$.

Let $v$ be the $N \times 1$ column vector of terminal rewards, such that, if the process ends in state $i$ then the terminal reward is the $i$th component of $v$.

For every $t \geq 0$, let $V^t(\pi, v)$ be the vector of total expected rewards earned during a process of duration $t$, using the policy $\pi$ and with terminal reward $v$. Then,

$$(2) \qquad V^t(\pi, v) = \int_0^t P(t,s;\pi) r(\pi(s)) ds + P(t;\pi)v \quad \text{for all} \quad t \geq 0.$$

## 2. Policies

A policy $\pi$ is called stationary and denoted by $f^\infty$ if $\pi(t) = f$, for all $t \geq 0$ and some $f \in F$. It is said to be initially stationary if there is a $0 \leq t < \infty$ and an $f \in F$ such that $\pi(s) = f$ for all $s \geq t$.

The policy $\pi^*$ is called an optimal policy if $V^t(\pi^*, v) \geq V^t(\pi, v)$ for all policies $\pi$ and all $t \geq 0$. An optimal policy maximizes total expected rewards in every component and for all $t \geq 0$ simultaneously.

A policy $\pi$ is <u>$\epsilon$-optimal</u> if, for $\epsilon > 0$,

$$\sup_{t \geq 0} \left| V^t(\pi^*, v) - V^t(\pi, v) \right|_\infty \leq \epsilon \ .$$

The policy $\pi$ is said to be piecewise constant and right continuous if the function $\pi(t)$ is piecewise constant in $t$ and if $\pi(t) = \lim_{s \to 0^+} \pi(t+s)$ for all $t \geq 0$.

The first theorem follows from Miller [14], and is given in Lembersky [12].

<u>Theorem 1.</u> Let $v \in R^N$ denote the vector of terminal rewards.

(i) There exists a piecewise constant, right continuous optimal policy.

If an optimal policy is restricted to be piecewise constant and right continuous, then

(ii) a necessary and sufficient condition for the piecewise constant right continuous policy, $\pi$, to be optimal is that

$$r(\pi(t)) + Q(\pi(t))V^t(\pi, v) \geq r(g) + Q(g)V^t(\pi, v)$$

for all $g \in F$ and $t \geq 0$, and further

(iii) for any optimal policy $\pi^*$, $V^t(\pi^*, v)$ is continuously differentiable in $t$, and

$$\frac{d}{dt} V^t(\pi^*, v) = r(\pi^*(t)) + Q(\pi^*(t))V^t(\pi^*, v) \quad \text{for all } t \geq 0.$$

It will be useful in the duration of this paper to restrict, without loss of generality, any optimal policy to be piecewise constant and right continuous.

To conclude this section, let $U = \lim_{t \to \infty} \frac{1}{t} V^t(\pi^*, v)$, which is well known to exist. Note that $U$ is the maximum possible long-run average return rate.

## 3. Decision Rules

Properties of decision rules and the related stationary policies will be discussed in this section and the results given will be used later sometimes without reference.

For a stationary policy $f^\infty$, the resulting process is the continuous time parameter stationary Markov process generated by the matrix $Q(f)$.

From (1), $P(t, s; f^\infty) = e^{(t-s)Q(f)}$ for all $t \geq s \geq 0$, where

$$e^{uQ(\cdot)} = I + \sum_{n=1}^{\infty} \frac{u^n}{n!} Q^n(\cdot) .$$

Note that $P(t, s; f^\infty) = P(t-s; f^\infty)$, for all $t \geq s \geq 0$. Also from Doob [4], for each $f \in F$, there exists an $N \times N$ stochastic matrix $P^*(f)$, such that $\lim_{t \to \infty} P(t; f^\infty) = P^*(f)$. Additionally,

$$P^*(f) = P^*(f)P^*(f) = P(t; f^\infty)P^*(f) = P^*(f)P(t; f^\infty),$$

for all $t \geq 0$, and

$$Q(f)P^*(f) = P^*(f)Q(f) = 0.$$

Next, define for each $f \in F$, $y(f) = \int_0^\infty [P(t; f^\infty) - P^*(f)]r(f)dt$. From [15], $y(f)$ exists and is the unique solution to the system

(3) $$P^*(f)y(f) = 0$$

(4) $$r(f) + Q(f)y(f) = P^*(f)r(f) .$$

For the next Lemma, parts (i) and (ii) may be found in [14], while (iii) is from [12].

<u>Lemma 1.</u> For $f \in F$ and $v \in R^N$,

(i) $V^t(f^\infty, v) = v + \sum_{n=1}^{\infty} \frac{t^n}{n!} Q^{n-1}(f)[r(f)+Q(f)v]$ for all $t \geq 0$ .

(ii) $\frac{d}{dt} V^t(f^\infty, v) = r(f) + Q(f)V^t(f^\infty, v)$ for all $t \geq 0$.

(iii) $V^t(f^\infty, v) - tP^*(f)r(f) \Rightarrow y(f) + P^*(f)v$ as $t \Rightarrow \infty$.

In view of Theorem 1 and Lemma 1 it may be useful to observe a different representation for $r(f) + Q(f)V^t(f^\infty, v)$ where $f \epsilon F$.

From Lemma 1, (i), with $v = 0$,

$$r(f) + Q(f)V^t(f^\infty, 0) = r(f) + Q(f)[t \cdot r(f) + \sum_{n=2}^{\infty} \frac{t^n}{n!} Q^{n-1}(f)r(f)]$$

$$= r(f) + Q(f)r(f)t + \sum_{n=2}^{\infty} \frac{t^n}{n!} Q^n(f)r(f)$$

$$= [I + \sum_{n=1}^{\infty} \frac{t^n}{n!} Q^n(f)]r(f) ,$$

hence

$$r(f) + Q(f)V^t(f^\infty, 0) = P(t; f^\infty)r(f) \tag{5}$$

For $v \neq 0$, it also follows using (2) that

$$r(f) + Q(f)V^t(f^\infty, v) = r(f) + Q(f)V^t(f^\infty, 0) + Q(f)P(t;f^\infty)v.$$

Then from (5),

$$r(f) + Q(f)V^t(f^\infty, v) = P(t;f^\infty)r(f) + Q(f)P(t;f^\infty)v ,$$

or

$$r(f) + Q(f)V^t(f^\infty, v) = P(t;f^\infty)[r(f)+Q(f)v] . \tag{6}$$

The equation in (6) gives an alternate way of finding $\frac{d}{dt} V^t(f^\infty, v)$ from $P(t;f^\infty)$ and is useful for later examples. Equations (5) and (6) also provided the motivation for the transformation given later in

Section 5 of this chapter.

Part (iii) of Lemma 1 implies that, for $f \in F$,

$$\lim_{t \to \infty} \frac{1}{t} V^t(f^\infty, v) = P^*(f)r(f) .$$

Hence define the set $F' = \{f \in F : P^*(f)r(f) = U\}$. Then for $f \in F'$, $f^\infty$ maximizes the long-run average return rate over all policies $\pi$. The set $F'$ is not empty as is shown in [15].

Define, for $f \in F$, $C(f)$ to be the set of recurrent states in the Markov process generated by $f^\infty$, and let $C = \bigcup_{f \in F'} C(f)$.

For $B$ a matrix, $B \succeq 0$ if the first non-zero element of each row of $B$ is positive, and $B \succ 0$ if $B \succeq 0$ and $B \neq 0$, where $0$ is a matrix of all zeroes. Also, if $A$ is of the same dimension as $B$, then $B \succeq (\succ) A$ if $B - A \succeq (\succ) 0$.

Define for $f$, $g \in F$, the vector

$$\psi(g, f; y(f)) = r(g) + Q(g)y(f) - P^*(f)r(f),$$

and the matrix

$$\Psi(g, f; y(f)) = (Q(g)P^*(f)r(f), \psi(g, f; y(f))) .$$

Let

$$G(f) = \{g \in f : \Psi(g, f; y(f)) \succ 0\} .$$

From [15], there exist decision rules $f \in F$ for which $G(f)$ is empty, and whenever $G(f)$ is empty, $f \in F'$ and $\Psi(g, f; y(f)) \preceq 0$

for every $g \in F$.

For $f \in F$, let $E(f) = \{g \in F: \psi(g, f; y(f)) = 0\}$. Since $f \in E(f)$, $E(f)$ is not empty. It should be noted that elements of $E(f)$ can be determined on a component by component basis. Also from [11], if $f \in F'$, then $E(f)$ is a subset of $F'$.

For $x \in R^N$, define

$$\begin{aligned} D(x) = \{f \in F: &(P^*(f)r(f), y(f) + P^*(f)x) \\ &\succeq (P^*(g)r(g), y(g) + P^*(g)x), \text{ for all } g \in F\} . \end{aligned}$$

It is known from Lanery [9] and Lembersky [11] that $D(x)$ is not empty. Also for $f \in D(x)$, let $x^* = y(f) + P^*(f)x$. Note that $D(x) \subset F'$.

## 4. Existence of $\epsilon$-Optimal Policies

The next two theorems are taken from [12] and will be used frequently in the following chapters.

Theorem 2. As $t \to \infty$, $V^t(\pi^*, v) - tU$ converges to some vector $V \in R^N$, and $V = V^*$.

Theorem 3. There is an $f \in F$ such that for every $\epsilon > 0$ there is a $t(\epsilon) > 0$ for which the initially stationary policy $\pi^\epsilon$,

$$\pi^{\epsilon}(t) = \begin{cases} \pi^*(t) & \text{for} \quad t < t(\epsilon) \\ f & \text{for} \quad t \geq t(\epsilon) \end{cases}$$

is $\epsilon$-optimal.

Following [12], let $F^*$ denote the set of decision rules, called preferred, from which the stationary parts of the policies $\pi^{\epsilon}$ of Theorem 3, are formed.

The last results are from [13], and the first one characterizes the set $F^*$ in terms of $V$.

Theorem 4. The following are equivalent.

(i) $f \in F^*$.

(ii) $f \in F'$ and $V = y(f) + P^*(f)V$.

(iii) $f \in F$, $Q(f)U = 0$, and $r(f) + Q(f)V = U$.

Lemma 2. If $f \in F$ and $Q(f)U = 0$, then $r(f) + Q(f)V \leq U$.

Theorem 5. There exists a $t^* \geq 0$ such that if $\pi^*(t) = f$ for any $t \geq t^*$, then $f \in F^*$. Further, either $\pi^*(t) = f$ for all $t \geq t^*$, or there are an infinite number of distinct intervals over which $\pi^*$ is constant and equal to $f$.

## 5. Transforming the Terminal Reward Vector

In this section an apparently useful transformation will be introduced for the rewards of a Markov decision chain. One of the results established will be that this transformation does not alter optimal policies.

For any $z \epsilon R^N$, define for each $a \epsilon A_i$, $i = 1, 2, \ldots, N$, the transformation

$$\bar{r}(i, a) = r(i, a) + \sum_{j=1}^{N} [q(j| i, a)z_j] .$$

Thus for any $(\cdot) \epsilon F$,

$$\bar{r}(\cdot) = r(\cdot) + Q(\cdot)z . \tag{7}$$

For any policy $\pi$ under which rewards are given by $\bar{r}(\cdot)$ as in (7), let $\bar{V}^t(\pi, w)$, denote the vector of total expected remaining rewards in a process $t$ time units from termination when the terminal reward is $w \epsilon R^N$.

Lemma 3. Let $v, w \epsilon R^N$. Set $z = v - w$ and let $\bar{r}(\cdot)$ be given by (7).

For any piecewise constant right continuous policy $\pi$,

$$V^t(\pi, v) - \bar{V}^t(\pi, w) = z \quad \text{for all} \quad t \geq 0 .$$

Proof: Assume

$$\pi(t) = \begin{cases} f_1, & 0 \le t < t_1 \\ f_2, & t_1 \le t < t_2 \\ \vdots & \vdots \\ f_k, & t_{k-1} \le t < t_k \\ \vdots & \vdots \end{cases} .$$

From Lemma 1, (i) it follows that

$$V^t(f^\infty, v) - \bar{V}^t(f^\infty, w) = z \quad \text{for all} \quad t \ge 0.$$

Therefore,

$$V^t(\pi, v) - \bar{V}^t(\pi, w) = z \quad \text{for all} \quad 0 \le t < t_1 .$$

So by induction assume that

$$V^t(\pi, v) - \bar{V}^t(\pi, w) = z \quad \text{for all} \quad 0 \le t < t_k .$$

Let $v_{t_k} = V^{t_k}(\pi, v)$ and $w_{t_k} = \bar{V}^{t_k}(\pi, w)$. By the continuity of $V^t(\pi, \cdot)$ and $\bar{V}^t(\pi, \cdot)$, $v_{t_k} - w_{t_k} = z$. Then from (2),

$$V^t(f^\infty_{k+1}, v_{t_k}) - \bar{V}^t(f^\infty_{k+1}, w_{t_k}) = V^t(f^\infty_{k+1}, 0) - \bar{V}^t(f^\infty_{k+1}, 0) + P(t; f^\infty_{k+1})z$$

for all $t \ge 0$. Now from Lemma 1, (i) again

$$V^t(f_{k+1}^\infty, 0) - \overline{V}^t(f_{k+1}^\infty, 0) = - \sum_{n=1}^{\infty} \frac{t^n}{n!} Q^n (f_{k+1})z$$

$$= z - [I + \sum_{n=1}^{\infty} \frac{t^n}{n!} Q^n(f_{k+1})]z$$

$$= z - e^{tQ(f_{k+1})} z$$

$$= z - P(t; f_{k+1}^\infty)z \quad \text{for all} \quad t \geq 0 \ .$$

Hence

$$V^t(f_{k+1}^\infty, v_{t_k}) - \overline{V}^t(f_{k+1}^\infty, w_{t_k}) = z \quad \text{for all} \quad t \geq 0.$$

Therefore, $V^t(\pi, v) - \overline{V}^t(\pi, w) = z$ for all $0 \leq t < t_{k+1}$, so the inductive argument is complete.

If the policy $\pi$ is initially stationary, the inductive argument above is modified accordingly.

For the remainder of this section consider the transformed system to have a terminal reward of $w \in R^N$ and rewards given by (7), with $z = v - w$.

Theorem 6. The piecewise constant right continuous policy $\pi*$ satisfies $V^t(\pi*, v) \geq V^t(\pi', v)$ for all policies $\pi'$ and all $t \geq 0$ if and only if $\overline{V}^t(\pi*, w) \geq \overline{V}^t(\pi', w)$ for all policies $\pi'$ and all $t \geq 0$; i.e., the transformation (7) does not change optimal policies.

Proof: From Lemma 3, it follows that

$$V^t(\pi^*, v) - V^t(\pi', v) = \overline{V}^t(\pi^*, w) - \overline{V}^t(\pi', w), \quad \text{for all} \quad t \geq 0,$$

from which the theorem follows.

Remark 1. Of interest is the case where $w = 0$ and $z = v - w = v$. In this instance the transformation (7) becomes $\overline{r}(\cdot) = r(\cdot) + Q(\cdot)v$ for all $(\cdot) \in F$. Then Theorem 6 implies that a Markov decision chain with a nonzero terminal reward may be transformed to one having a zero terminal reward, without altering optimal policies. Hence to prove results regarding $\pi^*$, it may be enough to prove these results when $v = 0$. For example, Miller's result stated as Theorem 1 was originally proven for $v = 0$. In view of the results in this section it follows that his result holds for any $v \in R^N$.

Let $\overline{U}$, $\overline{F}'$, $\overline{V}$, $\overline{F}^*$, $\overline{y}(\cdot)$, $D(\cdot)$, and $\overline{x}^*$ be the obvious analogs in the transformed system of respectively, $U$, $F'$, $V$, $F^*$, $y(\cdot)$, $D(\cdot)$ and $x^*$. The relationship between these quantities is given by the following corollary.

Corollary 1.

(i) $U = \overline{U}$.

(ii) $F' = \overline{F}'$.

(iii) $V = \overline{V} + z$.

(iv) $F^* = \overline{F}^*$.

(v) For any $f$, $y(f) = \bar{y}(f) + [I-P^*(f)]z$.

(vi) $D(v) = \bar{D}(w)$.

(vii) $v^* = \bar{w}^* + z$.

Proof: Since $P^*(\cdot)Q(\cdot) = 0$, for all $(\cdot) \in F$, it is clear that $U$ is unchanged by (7), hence $F'$ is also unchanged.

From Lemma 3 and Theorem 6, for any optimal policy $\pi^*$,

$$V^t(\pi^*, v) - tU = \bar{V}^t(\pi^*, w) - t\bar{U} + z \quad \text{for all} \quad t \geq 0.$$

Hence as $t \to \infty$, it follows that $V = \bar{V} + z$ which implies that for any $f \in F$,

$$\begin{aligned} r(f) + Q(f)V &= r(f) + Q(f)[\bar{V}+z] \\ &= \bar{r}(f) + Q(f)\bar{V}. \end{aligned}$$

Thus from Theorem 4, $F^* = \bar{F}^*$.

From Lemma 1, (iii), for any $f \in F$,

$$y(f) + P^*(f)v = \lim_{t \to \infty} [V^t(f^\infty, v) - tP^*(f)r(f)]$$

and

$$\bar{y}(f) + P^*(f)w = \lim_{t \to \infty} [\bar{V}^t(f^\infty, w) - tP^*(f)\bar{r}(f)].$$

So by Lemma 3 and (7),

$$\bar{y}(f) + P^*(f)w = \lim_{t\to\infty} [V^t(f^\infty, v) - z - tP^*(f)r(f)]$$

$$= y(f) + P^*(f)v - z .$$

Thus $D(v) = \bar{D}(w)$ and $\bar{w}^* = v^* - z$.

Remark 2. Consider Corollary 1, (vi) with $w = 0$. Then $D(v) = \bar{D}(0)$. In view of the algorithm given by Veinott [17] for finding a decision rule in $D(0)$, it is clear that his algorithm may also be used for finding a decision rule in $D(v)$ by first applying transformation (7) with $z = v$ to the system. Hence the above results provide an alternate to the algorithms of Lanery [9] and Teghem [16] for finding a decision rule in $D(v)$.

# II. OPTIMAL POLICIES

## 0. Introduction

Section 1 considers the problem of continuous time decision processes with $n \geq 1$ absorbing states. It will be established that when $n > 1$ there need not exist a stationary optimal policy, however, the set $F^*$ will be shown to be equal to the set $D(v)$ hence giving the entire set $F^*$. An application of this result to a control problem is also included.

Single class decision rules are utilized in Section 2 to give a characterization of the set $F^*$ in terms of the set $E(f)$, where $f \in F$ has $G(f)$ empty and each $g \in E(f)$ is single class.

In Section 3 some counterexamples are given. The motivation for these examples comes both from possible conjectures stated by other authors and from conjectures related to results given here.

Section 4 of this chapter gives necessary and sufficient conditions for a stationary policy to attain maximal expected rewards on various recurrent states. When the recurrent states include the entire set $S$ this results in necessary and sufficient conditions for a stationary policy to be optimal. These results may also be extended to give necessary and sufficient conditions for initially stationary policies to attain maximal expected rewards on various recurrent states when total expected rewards using $\pi^*$ are known at some time point $\bar{t} \geq 0$. Additional implications of these results suggest a natural refinement of the set $F^*$ when stationary optimal policies are known to exist.

The final section of this paper discusses the question of existence of initially stationary optimal policies and, for $N = 3$, presents a possible approach to solving this problem.

## 1. Absorbing States

Several authors have treated the case of a single absorbing (stopping) state, (see Veinott [18]) and have established the existence of a stationary optimal policy. This section considers continuous time decision processes with $n \geq 1$ absorbing states.

Consider a Markov decision chain with the following structure. Let

$$S = \{1, 2, \ldots, n, n+1, \ldots, N\},$$

where $C = \{1, 2, \ldots, n\}$ are absorbing states under every decision rule, and $T = \{n+1, \ldots, N\}$ are transient under every decision rule. Thus for each $i \in C$, $A_i$ consists of a single action $a^i$ for which $q(j | i, a^i) = 0$ for all $j \in S$. Further, for any $f \in F$, $C(f) = C$. The following notation will be used, with the obvious partitions for recurrent and transient states. Write

$$Q(f) = \begin{bmatrix} 0_{n \times n} & 0 \\ Q_{TR}(f) & Q_T(f) \end{bmatrix} , \quad P^*(f) = \begin{bmatrix} I_{n \times n} & 0 \\ P^*_{TR}(f) & 0 \end{bmatrix} .$$

It is clear that for $i \epsilon C$, $U_i = r(i, a^i)$ and, from (3), I, that $y_i(\cdot) = 0$. In other words,

$$y(f) = \begin{bmatrix} 0_{n \times 1} \\ y_T(f) \end{bmatrix} , \quad \text{and} \quad r(f) = \begin{bmatrix} U_C \\ r_T(f) \end{bmatrix} .$$

If $v = \begin{bmatrix} v_C \\ v_T \end{bmatrix}$ is the vector of terminal rewards, then by Theorem 2, I and since for all $\pi$,

$$V_i^t(\pi, v) = v_i + t \cdot r(i, a^i) \quad \text{for each} \quad i \epsilon C ,$$

$$V = \begin{bmatrix} V_C \\ V_T \end{bmatrix} = \begin{bmatrix} v_c \\ V_T \end{bmatrix} .$$

Theorem 1.

$$F^* = D(v) .$$

Proof: From Theorem 4, I, $f \epsilon F^*$ if and only if $f \epsilon F'$ and $V = y(f) + P^*(f)V$.

Hence, for any $f \epsilon F^*$,

$$V_T = y_T(f) + (P^*_{TR}(f)\ 0)V = y_T(f) + P^*_{TR}(f)V_C = y_T(f) + P^*_{TR}(f)v_C .$$

Now since $V^t(\pi^*, v) \geq V^t(f^\infty, v)$ for all $t \geq 0$, it follows from Theorem 2, I and Lemma 1, (iii), I, that

$$y_T(f) + P^*_{TR}(f)V_C \geq y_T(\cdot) + P^*_{TR}(\cdot)v_C \quad \text{for all } (\cdot) \in F' .$$

Thus $f \in D(v)$.

Further, for any other $g \in D(v)$,

$$V_T = y_T(f) + P^*_{TR}(f)v_C = y_T(g) + P^*_{TR}(g)v_C = y_T(g) + P^*_{TR}V_C ,$$

so $g \in F^*$. Hence $F^* = D(v)$.

Remarks.

1. The vector $V$ is given by

$$V = v^* = \begin{bmatrix} v_c \\ y_T(f) + P^*_{TR}(f)v_c \end{bmatrix}, \quad \text{for any } f \in D(v).$$

2. The set $F^*$ and the vector $V$ are independent of $v_T$.

3. Theorem 1 gives the entire set $F^*$ and from Theorem 5, I, in the case where $D(v)$ is a singleton set, provides the rule that must be used as the initially stationary piece of every optimal policy. However, when $n \geq 2$ the rule (or rules) provided need not be stationary optimal (i.e., in general there need not exist any stationary optimal policy, as is always the case when $n = 1$). This may be seen

from the next example.

Counterexample 1. Assume there are two decision rules, $F = \{f, g\}$, differing only in their actions and rewards in the fourth state. Let

$$r(f) = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \quad Q(f) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & -1 \end{bmatrix};$$

$$r(g) = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2.5 \end{bmatrix}, \quad Q(g) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & -1 \end{bmatrix}.$$

Set $v = 0$. Then by Theorem 1,I there exists a $t' > 0$ such that $\pi^*(t) = g$, for $0 \leq t < t'$.

Note that

$$P^*(f) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix} \quad \text{and} \quad P^*(g) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

so $F' = \{f, g\}$. Using (3), I, and (4), I, it follows that

$$y_T(f) + P_T^*(f)v_c = \binom{-2}{1},$$

while

$$y_T(g) + P_T^*(g)v_c = \binom{-2}{.5}.$$

Hence $F^* = \{f\}$. Therefore, neither $f^\infty$ nor $g^\infty$ is optimal.

Applications. As discussed at the beginning of this section the case of a single absorbing state has received much attention and gives a useful model for many problem areas. To indicate potential uses for the structure with several absorbing states, applications to environmental control problems will be briefly indicated.

A recent survey by Jaquette [7] provides a good summary of the literature in the area of control problems for biological populations, including the usage of discrete time Markov decision chains and optimal stopping rules.

Becker [2] develops an analytical model for the control of pests in a habitat. He assumes the growth of a pest population to be according to a simple birth, death, and immigration process with the control of the growth taken over a finite time interval. Becker obtains functional characteristics of the optimal control functions $\rho(t) = \lambda(t) - \mu(t)$ which indicates an action to be taken causing an impact on birth rate, death rate, or on both. Control actions using insecticides, parasites, predators or possibly the introduction of diseased pests to spread a virus into the population would then be considered. Becker then uses the calculus of variations to find an a priori optimal $\rho(t)$.

An alternative formulation for the control of pests as a continuous time Markov decision chain with absorbing states is given next and the preceding results of this section may be applied to obtain decision rules in the set $F^*$.

Consider a population of pests in a given habitat. When the level of pests can become damaging to the crop or renewable resource in the habitat, the problem of controlling the pests is of interest. It becomes necessary to weigh the cost of controlling the pests against the damage done by them. A possible formulation in the framework of this paper involves identifying the appropriate states of the system, available actions in each state, transition rates for each action, corresponding reward rates and the vector of terminal rewards.

Assume that measurements can be taken to determine the degree of infestation to a crop in a given area and the residual level of the treatment (e.g., D.D.T.) in the related environment. The process is then assumed to be observed in one of two classes of states, one class consisting of absorbing states, and the other consisting of transient states. The various absorbing states represent both desirable and undesirable permanent situations, while the transient states are such that the system is eventually absorbed into one of the absorbing states. If no action is taken in a transient state, it is then possible for the system to move to an undesirable absorbing state (the crop is destroyed). If an action is taken in a transient state, it is expected that the system will move to either a desirable absorbing state (the crop is in good harvest condition) or an undesirable absorbing state (residual level of treatment is too high) with varying rates depending on the action taken.

Define as follows:

State 1. Less than $K_1$ pests per unit of area.

Here $K_1$ represents an acceptable level of pests (i.e., $K_1$ may be the threshold where losses are considered negligible and the birth and death rates are such that the population is assumed to stay less than $K_1$).

Let $A_1 = \{NA\}$ indicating no treatment action is to be taken, and $q(j|1,NA) = 0$ for all $j \in S$. Also, $r(1,NA) = 0$, no rewards or costs are earned until the crop is harvested at $t = 0$. Set $v_1 = M_1 > 0$, where $M_1$ is the market value of an undamaged crop at harvest.

State 2. Residual level of treatments applied is too high and damage is done to the environment.

Let $A_2 = \{NA\}$ and $q(j|2,NA) = 0$ for $j \in S$. Set $v_2 = -M_2 < -M_1$ where the damage done to the environment is considered to be monetarily greater than the value of the crop. Also let $r(2,NA) = 0$.

State 3. More than $K_2$ pests per unit of area $(K_2 > K_1)$. This state indicates that the crop is destroyed.

Let $A_3 = \{NA\}$ and $q(j|3,NA) = 0$ for all $j \in S$. Set $v_3 = 0$ and $r(3,NA) = 0$.

Let $K_1, K_1+1, \ldots, K_2$ indicate the number of pests per unit of area and set $S = \{1,2,3,K_1,K_1+1,\ldots,K_2\}$. For $i = K_1,\ldots,K_2$ let $A_i = \{NA, a_1,\ldots,a_{k'}\}$, where $a_1,\ldots,a_{k'}$ are levels of treatment with $a_j < a_{j+1}$, $j = 1,\ldots,k'-1$, and NA indicates no treatment.

For $i \in \{K_1+1,\ldots,K_2-1\}$, let $q(i+1|i,NA) = \lambda_i$, $q(i-1|i,NA) = \mu_i$ and $q(i|i,NA) = -(\mu_i+\lambda_i)$ with $\lambda_i > \mu_i$. For $i = K_1$, let $q(K_1+1|K_1,NA) = \lambda_{K_1}$, $q(1|K_1,NA) = \mu_{K_1}$ and $q(K_1|K_1,NA) = -(\lambda_{K_1}+\mu_{K_1})$. For $i = K_2$, let $q(3|K_2,NA) = \lambda_{K_2}$, $q(K_2-1|K_2,NA) = \mu_{K_2}$ and $q(K_2|K_2,NA) = -(\lambda_{K_2}+\mu_{K_2})$.

For $a_j$, $j = 1,\ldots,k'$, let $q(1|i,a_j) = L_{ij1}$, $q(2|i,a_j) = L_{ij2}$, and $q(2|i,a_j) = -(L_{ij1}+L_{ij2})$ for $i = K_1,\ldots,K_2$, where $L_{ij1}$ increases as a function of $j$ for each $i$ and $L_{ij2}$ increases as a function of $j$ for each $i$.

Set $v_i = M_1 - di$, $i = K_1, \ldots, K_2$, where $d > 0$ and $di$ represents the dollar loss of crop at harvest due to the level of pests. Also, let $r(i, a_j) < 0$ and strictly decreasing in $j$, with $-r(i, a_j)$ representing the cost of the treatment.

Hence the process described has $C = \{1, 2, 3\}$ and $T = \{K_1, K_1+1, \ldots, K_2\}$. Theorem 1 may then be used to find an $f \in F^*$.

The number of computations in finding such an $f \in F^*$ can be quite large for any problem involving several transient states. A considerably simplified, but illustrative variation of this form is given next having only one transient state, with several actions available in this transient state.

Consider states 1 and 2 to be as given before. Let state 3 denote $K_1$ or more pests per unit of area and $S = \{1, 2, 3\}$. In state 3 a treatment will necessarily be taken, the optimal level of treatment to be applied as a function of time is to be determined. Thus assume that $A_3 = \{a, 2a, 3a, \ldots, ka\}$, where $a, \ldots, ka$ are levels of treatment with $ja < (j+1)a$, $j = 1, \ldots, k-1$. Further, assume that

$$q(1|3, ja) = K + \alpha j$$

$$q(2|3, ja) = e^j ,$$

i.e., the transition rate to state 1 increases linearly as a function of $j$ while the transition rate to state 2 increases exponentially in $j$. Further assume that $K > 0$, $\alpha > 0$ and that $K+\alpha > e$. Set $v_3 = 0$ and let $r(3, j\alpha) = -C - \beta j$, $C > 0$, $\beta > 0$, so that the costs increase as a linear function of $j$.

Now for each $(\cdot) \in F$ with action $j$ in state 3, $1 \leq j \leq k$,

$$Q(\cdot) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ K+\alpha j & e^j & -K-\alpha j-e^j \end{bmatrix}, \quad r(\cdot) = \begin{bmatrix} 0 \\ 0 \\ -C-\beta j \end{bmatrix}$$

$$P^*(\cdot) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{K+\alpha j}{K+\alpha j+e^j} & \frac{e^j}{K+\alpha j+e^j} & 0 \end{bmatrix}, \quad y(\cdot) = -\begin{bmatrix} 0 \\ 0 \\ \frac{C+\beta j}{K+\alpha j+e^j} \end{bmatrix}$$

Hence

$$y(\cdot) + P^*(\cdot)v = \begin{bmatrix} M_1 \\ -M_2 \\ \frac{-(C+\beta j)}{K+\alpha j+e^j} + \frac{M_1(K+\alpha j)}{K+\alpha j+e^j} - \frac{M_2 e^j}{K+\alpha j+e^j} \end{bmatrix}$$

Select $j^* \in \{1, \ldots, k\}$ such that

$$\frac{-C-\beta j^* + M_1(K+\alpha j^*) - M_2 e^{j^*}}{K+\alpha j^*+e^{j^*}} \geq \frac{-C-\beta j + M_1(K+\alpha j) - M_2 e^j}{K+\alpha j+e^j}$$

for all $j = 1, 2, \ldots, k$ and choose $f \in F$ with $f(3) = j^*a$. By Theorem 1, $f \in F^*$. In other words, a preferred rule may be easily computed.

By using the fact that Theorem 1 implies $V = v^*$, under a condition on the problem constants, it will be established that $f^\infty$ is, in fact, an optimal policy. To see this, first note that for $g \in F$ with $g(3) = ja$, $1 \leq j \leq k$, it follows by induction that for all $n \geq 1$,

$$Q^n(g) = [-(K+\alpha j+e^j)]^{n-1} Q(g).$$

Therefore, by Lemma 1, (i), I, for all $t \geq 0$,

$$\begin{aligned} V_3^t(g^\infty, v) &= t(-C-\beta j+(K+\alpha j)M_1+e^j(-M_2)) \\ &\quad + \sum_{n=2}^{\infty} \frac{t^n}{n!}((K+\alpha j)(-K-\alpha j-e^j)^{n-2} e^j(-K-\alpha j-e^j)^{n-2}(-K-\alpha j-e^j)^{n-1}) \\ &\quad \times \begin{bmatrix} 0 \\ 0 \\ -C-\beta j+(K+\alpha j)M_1+e^j(-M_2) \end{bmatrix} \\ &= \sum_{n=1}^{\infty} \frac{t^n}{n!}[-C-\beta j+M_1(K+\alpha j)-M_2 e^j](-K-\alpha j-e^j)^{n-1} \quad = \end{aligned}$$

$$= \frac{C+\beta j-M_1(K+\alpha j)+M_2e^j}{K+\alpha j+e^j} \sum_{n=1}^{\infty} \frac{[-t(K+\alpha j+e^j]^n}{n!}$$

$$= \frac{C+\beta j-M_1(K+\alpha j)+M_2e^j}{K+\alpha j+e^j} [e^{-t(K+\alpha j+e^j)} -1].$$

Hence

$$V^t(g^\infty, v) = \begin{bmatrix} M_1 \\ -M_2 \\ V_3^t(g^\infty, v) \end{bmatrix} \quad \text{for all} \quad t \geq 0 .$$

Note that $P^*(\cdot)r(\cdot) = 0$ for all $(\cdot) \in F$ so that $F = F'$ and $U = 0$.

For all $(\cdot) \in F$ and all $t \geq 0$, let

$$\phi^t(f, \cdot; v) = r(f) + Q(f)V^t(f^\infty, v) - r(\cdot) - Q(\cdot)V^t(f^\infty, v) .$$

By Lemma 1, (iii), I, as $t \rightarrow \infty$, $\phi^t(f, \cdot; x) \rightarrow -r(\cdot) - Q(\cdot)[y(f)+P^*(f)v]$. Now since $V = v^*$ and $f \in D(v)$, $y(f) + P^*(f)v = V$. Hence, by Lemma 2, I, $\phi^t(f, \cdot; x) \rightarrow -r(\cdot) - Q(\cdot)V$ as $t \rightarrow \infty$, and $-r(\cdot) - Q(\cdot)V \geq 0$ for all $(\cdot) \in F$. Further, for all $g \in F$ such that $g(3) = ja$,

$$\phi_3^t(f,g;x) = \beta(j-j^*) - \alpha M_1(j-j^*) + M_2(e^j - e^{j^*})$$

$$+ [\alpha(j-j^*)+e^j-e^{j^*}]\left[\frac{C+\beta j^*-M_1(K+\alpha j^*)+M_2e^{j^*}}{K+\alpha j^*+e^{j^*}}\right]$$

$$\times [e^{-t(K+\alpha j^*+e^{j^*})} - 1] .$$

Therefore, $\phi^t(f,g;x)$ is of the form

$$\phi^t(f,g;x) = \begin{bmatrix} 0 \\ 0 \\ a+b[e^{-\gamma t}-1] \end{bmatrix},$$

where $a$ and $b$ depend on $g(3)$, but not $t$.

Lemma 1. If $r(f) + Q(f)v \geq r(\cdot) + Q(\cdot)v$ for all $(\cdot) \epsilon F$, then $f^\infty$ is an optimal policy.

Proof: Since for any $g \epsilon F$, $r(f) + Q(f)v \geq r(g) + Q(g)v$ and $\phi_3^t(f,g;v) = a + b[e^{-\gamma t}-1]$, it follows that $a \geq 0$. Since $\lim_{t\to\infty} \phi^t(f,g;v) \geq 0$ and since $\phi^t(f,g;v)$ is monotone for all $t \geq 0$, $\phi^t(f,g;v) \geq 0$ for all $t \geq 0$. From Theorem 1, I, $f^\infty$ is an optimal policy.

Theorem 2. If $\frac{C+\beta+M_2e-M_1K}{M_1} \leq \alpha < \frac{\beta+M_2e(e-1)}{M_1}$, then $j^* = 1$ and the policy $f^\infty$ is optimal, where $f(3) = j^*$.

Proof: For $\alpha \leq \frac{\beta+M_2e(e-1)}{M_1}$, it follows that for all $j \geq 2$, $\alpha M_1(j-1) \leq \beta(j-1) + M_2(j-1)(e^2-e)$, and, by induction on $j \geq 2$, $(j-1)(e^2-e) \leq e^j - e$, so $\alpha M_1(j-1) \leq \beta(j-1) + M_2(e^j-e)$. From this it follows that

$$-C - \beta + M_1(K+\alpha) - M_2e \geq -C - \beta j + M_1(K+\alpha j) - M_2e^j ,$$

for $j \geq 2$.

Since $\alpha \geq \frac{C+\beta+M_2e-M_1K}{M_1}$, it follow that $-C - \beta + M_1(K+\alpha) - M_2e \geq 0$. Hence for all $j \geq 2$,

$$\frac{-C-\beta+M_1(K+\alpha)-M_2e}{K+\alpha+e} \geq \frac{-C-\beta j+M_1(K+\alpha j)-M_2e^j}{K+\alpha j+e^j} ,$$

so $j^* = 1$. Also, for $j \geq 2$

$$-C - \beta + M_1(K+\alpha) - M_2e \geq -C - \beta j + M_1(K+\alpha j) - M_2e^j$$

implies that $r(f) + Q(f)v \geq r(\cdot) + Q(\cdot)v$ for all $(\cdot) \in F$. Hence from Lemma 1, $f^\infty$ is optimal.

## 2. Single Class Decision Rules and the Set F*

For $f \in F$, write

$$r(f) = \begin{bmatrix} r_R(f) \\ r_T(f) \end{bmatrix} \quad \text{and} \quad y(f) = \begin{bmatrix} y_R(f) \\ y_T(f) \end{bmatrix} \quad ;$$

the partition indicating the recurrent and transient states for the Markov chain determined by $f$. The next Lemma will be needed for the main result of this section.

Lemma 2. Let $h, g \epsilon F'$ and assume $h(i) = g(i)$ for all $i \epsilon C$. Then $y_R(h) = y_R(g)$ .

Proof: It is clear that $h$ and $g$ have the same recurrent classes, $r_R(h) = r_R(g) \equiv r_R$, and upon partitioning $Q(h)$ and $Q(g)$ corresponding to the recurrent states, $C(h) = C(g)$, and the transient states so that

$$Q(h) = \begin{bmatrix} Q_R & 0 \\ Q_{TR}(h) & Q_T(h) \end{bmatrix}, \quad Q(g) = \begin{bmatrix} Q_R & 0 \\ Q_{TR}(g) & Q_T(g) \end{bmatrix},$$

it follows from standard Markov chain theory that

$$P^*(h) = \begin{bmatrix} P^*_R & 0 \\ P^*_{TR}(h) & 0 \end{bmatrix} \quad \text{and} \quad P^*(g) = \begin{bmatrix} P^*_R & 0 \\ P^*_{TR}(g) & 0 \end{bmatrix} .$$

Then from (3), I and (4), I, $y_R(h)$ and $y_R(g)$ are unique solutions to the same set of equations.

$$\begin{cases} P^*_R y = 0 , \\ r_R + Q_R y_R = P^*_R r_R , \end{cases}$$

hence $y_R(h) = y_R(g)$.

The next Lemma is from Lembersky [13].

Lemma 3. If $f \in F$ and $G(f)$ is empty, then there is a $g \in F^*$;

(i) such that $g(i) = f(i)$ for all $i \in C(f)$, and

(ii) there is an $h \in E(f)$ such that $h(i) = g(i)$ for all $i \in C$.

The following result strengthenes an earlier theorem by Lembersky in [11] with the hypothesis that each $g \in F'$ be single class. A decision rule $g$ is single class when the resulting Markov chain has a single recurrent class (and possibly some transient states).

Theorem 3. If $f \in F$ with $G(f)$ empty and each $g \in E(f)$ is single class, then $E(f) = F^*$.

Proof: Let $g \in F^*$, $h \in E(f)$ such that

$$h(i) = \begin{cases} g(i), & i \in C \\ f(i), & i \notin C . \end{cases}$$

Such a choice is possible from Lemma 3, (ii). Then since $E(f) \subset F'$ and by Theorem 4, I, both $h$ and $g$ are in $F'$. Therefore, $C(g) = C(h)$. Clearly $h$ is single class, hence $g$ is single class. From Markov chain theory, when a decision rule is single class, each row of $P^*(\cdot)$ is the same. Hence there exists a scalar $u$, such that $U = u \circ 1$, where $u \circ 1$ denotes the $N \times 1$ column vector each of whose components is equal to $u$. Further, there exists a scalar $k$ such that $P^*(g)V = k \circ 1$. By Theorem 4, I, this implies $V = y(g) + k \circ 1$, and thus for any $(\cdot) \in F$, $Q(\cdot)V = Q(\cdot)y(g)$ and $Q(\cdot)U = 0$, so from Lemma 2, I, $G(g) = \phi$.

Since $h \in E(f)$, by (4), I, $Q(h)y(f) = Q(h)y(h)$, or $Q(h)[y(h)-y(f)] = 0$. Also, by (3), I, $P^*(h)[y(h)-y(f)] = P^*(h)[-y(f)]$. So from Miller [15, p. 567], $y(h) - y(f) = P^*(h)[-y(f)]$, and since $h$ is single class, there is a scalar $\ell$ such that

$$y(h) - y(f) = \ell \circ 1. \tag{1}$$

It follows that $G(h)$ is empty.

From Lemma 2, $y_R(h) = y_R(g)$. Also, $G(g)$ empty and $Q(h)U = 0$ implies

$$r(h) + Q(h)y(g) - U \leq 0 ,$$

and since

$$r(h) + Q(h)y(h) - U = 0 ,$$

$$Q(h)[y(g)-y(h)] \leq 0 . \tag{2}$$

Now using the notation from the proof of Lemma 2, (2) gives

$$Q_{TR}(h)[y_R(g)-y_R(h)] + Q_T(h)[y_T(g)-y_T(h)] \leq 0 ,$$

or

$$Q_T(h)[y_T(g)-y_T(h)] \leq 0 .$$

Now since $Q_T(h)$ has an inverse with all nonpositive elements, it follows that $y_T(g) \geq y_T(h)$.

Similarly, $G(h)$ empty gives $y_T(g) \leq y_T(h)$, so $y(g) = y(h)$. From (1), then

$$y(f) = y(g) - \ell \circ 1 . \tag{3}$$

Let $h' \epsilon E(f)$. From (1) and (3), $r(h') + Q(h')y(g) - U = 0$, and since $V = y(g) + k \circ 1$, it follows that $r(h') + Q(h')V - U = 0$ and $Q(h')U = 0$. So from Theorem 4, I, $h' \epsilon F^*$, and $E(f) \subset F^*$.

Let $h' \epsilon F^*$. Then $Q(h')U = 0$, and $r(h') + Q(h')V - U = 0$. From (3) and $V = y(g) + k \circ 1$, $h' \epsilon E(f)$, so $F^* \subset E(f)$ and the proof is complete.

The first corollary follows immediately from the argument that $G(g)$ is empty that is given in the proof of the preceding Theorem.

<u>Corollary 1.</u> If $f \epsilon F^*$ and $f$ is single class, then $G(f)$ is empty.

The next corollary offers an alternative hypothesis under which $E(f) = F^*$ for an $f$ in $F$ such that $G(f)$ is empty.

Corollary 2. If $f \in F$ with $G(f)$ empty, $f$ is single class, and $C(f) = C$, then $E(f) = F^*$.

Proof: From Lemma 3, (i), there exists a $g \in F^*$ such that $f(i) = g(i)$ for all $i \in C$. The result then follows using the arguments of the proof of Theorem 3.

## 3. Counterexamples

In this section several examples are given. Motivated by Theorem 3, a possible assertion might be that for some $f$ in $F$ with $G(f)$ empty, $E(f) \cap F^*$ is always non-empty. The first example of this section will show that this need not be true. (It is also apparent from Theorem 1 that in general the set $F^*$ is not independent of the terminal reward $v$, as it is under the hypothesis of Theorem 3 or Corollary 2. ).

Counterexample 2. There are two decision rules, $F = \{f, g\}$, differing only in the third state. Let

$$Q(f) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3/4 & 1/4 & -1 \end{bmatrix}, \quad r(f) = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}; \quad Q(g) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}, \quad r(g) = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}.$$

Then

$$P^*(f) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3/4 & 1/4 & 0 \end{bmatrix}, \quad y(f) = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix},$$

$$P^*(g) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad y(g) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ and } U = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Note that

$$r(g) + Q(g)y(f) - U = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix};$$

hence $G(f)$ is empty. Set

$$v = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}.$$

Then from Theorem 1, it follows that $F^* = \{g\}$. Now $G(f)$ is empty, $g \notin E(f)$, so $E(f) \cap F^*$ is empty. Further, since

$$r(f) + Q(f)y(g) - U = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

$f$ is the only decision rule with $G(\cdot)$ empty.

Several interesting conclusions regarding the set $F^*$ will be drawn from the next example.

Counterexample 3. Again there are two decision rules, $F = \{f, g\}$, differing in the second state. Let

$$Q(f) = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad r(f) = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 1 \end{bmatrix};$$

$$Q(g) = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad r(g) = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then

$$P^*(f) = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}, \quad P^*(g) = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}, \quad U = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Set $v = 0$. Then by Theorem 1, I, $\pi^*(t) = f$ on $0 \leq t < t^1$, for some $t^1 > 0$. To see if $t^1$ may be arbitrarily large, compute $V^t(f^\infty) \equiv V^t(f^\infty, 0)$. By induction,

$$Q^n(f) = (-1)^n \begin{bmatrix} 2^{n-1} & -2^{n-1} & 0 & 0 \\ -2^{n-1} & 2^{n-1} & 0 & 0 \\ 2^{n-1}-1 & -2^{n-1} & 1 & 0 \\ 2^{n-1}-1 & -2^{n-1} & 0 & 1 \end{bmatrix} ,$$

so by Lemma 1, (i), I,

$$V^t(f^\infty) = t \begin{bmatrix} 0 \\ 2 \\ 1 \\ 1 \end{bmatrix} + \sum_{n=2}^{\infty} \frac{t^n}{n!} (-1)^{n-1} \begin{bmatrix} -2^{n-1} \\ 2^{n-1} \\ -2^{n-1}+1 \\ -2^{n-1}+1 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2} + t + \frac{1}{2} e^{-2t} \\ \frac{1}{2} + t - \frac{1}{2} e^{-2t} \\ \frac{1}{2} + t + \frac{1}{2} e^{-2t} - e^{-t} \\ \frac{1}{2} + t + \frac{1}{2} e^{-2t} - e^{-t} \end{bmatrix} \quad \text{for all} \quad t \geq 0 .$$

Then

$$r(f) + Q(f)V^t(f^\infty) - r(g) - Q(g)V^t(f^\infty) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -1+e^{-t} \\ 0 \\ 0 \end{bmatrix}$$

$$\text{for all} \quad t \geq 0 .$$

Therefore, since $e^{-t} > 0$ for all $t > 0$, it follows from

Theorem 1, I, that $\pi^*(t) = f$ for all $t \geq 0$, and that $f^\infty$ is the unique optimal policy. Also

$$\lim_{t \to \infty} [V^t(f^\infty) - tU] = \begin{bmatrix} -1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = V.$$

So, $r(g) + Q(g)V - U = 0$, and thus by Theorem 4, I $F^* = \{f, g\}$.

The following observations may be made from this example:

1. The unique optimal policy is stationary, yet the set $F^*$ has two elements. Note that $g \in F^*$, but $\pi^*(t) \neq g$ for any $t \geq 0$. Hence the converse of the first part of Theorem 5, I, need not be true.
2. The hypotheses in Corollary 2 imply the existence of an $f \in F^*$ such that $C(f) = C$. Note that in this example such a decision rule does not exist.
3. By Corollary 1, the set $G(g)$ is empty. Also,

$$y(g) = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \text{so} \quad E(g) = \{f, g\} .$$

> Thus, if the algorithm given by Lembersky in [13] for computing a decision rule h as in Lemma 3, (ii), starts with g, while giving a rule in F*, it fails to give the decision rule f that in fact forms the stationary optimal policy.

From Theorem 3, I, for every $\epsilon > 0$ there exists an initially stationary $\epsilon$-optimal policy. The obvious question then, is whether there always exists an initially stationary optimal policy. For $N = 2$, an affirmative answer may be given from Miller [14, Section 6], when there is at least one ergodic decision rule f. The assumption of an ergodic decision rule is unnecessary, as is shown in Lembersky [11, Section 4. 7]. In this section a five state example is given which shows that there need not always exist an initially stationary optimal policy.

A brief statement of the idea behind the example is given now. In state five, there are two actions available, the choice of which allows the system to move to one of two disjoint classes of states. For one of these classes, total rewards earned increase at a constant rate, while in the other, the derivative of the total rewards earned over time oscillates above and below this constant rate. It is then reasonable to expect the optimal policy to also oscillate in its choice of action for state five.

Counterexample 4. There are two decision rules, $F = \{f, g\}$, which differ only in the fifth state. Let

$$Q(f) = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad r(f) = \begin{bmatrix} 6 \\ 3 \\ 3 \\ 4 \\ 4 \end{bmatrix},$$

$$Q(g) = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad r(g) = \begin{bmatrix} 6 \\ 3 \\ 3 \\ 4 \\ 3 \end{bmatrix}.$$

Set $v = 0$, $V^t(\pi) = V^t(\pi, 0)$, and note that for any policy $\pi$,

$$r(f) + Q(f)V^t(\pi) - [r(g)+Q(g)V^t(\pi)] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1+V^t_4(\pi)-V^t_1(\pi) \end{bmatrix}$$

for all $t \geq 0$. Then for any $\pi$ and all $i \neq 5$, $V^t_i(\pi) = V^t_i(f^\infty) = V^t_i(g^\infty)$. So, by letting $\delta(t) = 1 + V^t_4(f^\infty) - V^t_1(f^\infty)$, it follows from Theorem 1, I, that for every optimal policy $\pi^*$,

$$\pi^*(t) = f, \quad \text{when} \quad \delta(t) > 0$$
$$= g, \quad \text{when} \quad \delta(t) < 0 \;.$$

To show that there is no initially stationary optimal policy, it will suffice to show that $\delta(t)$ continually oscillates above and below zero as a function of $t$. To accomplish this it will be necessary to find $V^t(f^\infty) \equiv V^t(f^\infty, 0)$ for all $t \geq 0$. The first step in calculating $V^t(f^\infty)$ is to find $P(t; f^\infty)$. The procedure used is essentially that of Karlin [8, p. 208].

First determine the eigenvalues $\lambda_1, \ldots, \lambda_N$ of $Q(f)$ and a complete system of associated right eigenvectors $w^{(1)}, \ldots, w^{(N)}$. Then $P(t;f^\infty) = W\Lambda(t)W^{-1}$, where $W$ is the matrix whose column vectors are, respectively $w^{(1)}, \ldots, w^{(N)}$ and $\Lambda(t) = \text{diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_N t})$.

The eigenvalues for $Q(f)$ are

$$\frac{-3 - \sqrt{3}\,i}{2}, \quad \frac{-3 + \sqrt{3}\,i}{2}, \quad 0, \quad 0, \quad -1 \;.$$

Let

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} .$$

Then

$$Q(f) = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} ,$$

so that if $x$ is an eigenvector for $A$ then $\binom{x}{0}$ is an eigenvector for $Q(f)$. Also $y$ an eigenvector for $B$ implies $\binom{0}{y}$ is an eigenvector for $Q(f)$. The eigenvalues for $A$ are distinct and those for $B$ are also distinct, hence $A$ and $B$ are necessarily diagonalizable and $Q(f)$ must then be diagonalizable.

It is easy to see that matrices of right eigenvectors for $A$ and $B$ are, respectively,

$$W_A = \begin{bmatrix} \frac{-1-\sqrt{3}\,i}{2} & \frac{-1+\sqrt{3}\,i}{2} & 1 \\ \frac{-1+\sqrt{3}\,i}{2} & \frac{-1-\sqrt{3}\,i}{2} & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad W_B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

The determinant of both $W_A$ and $W_B$ is nonzero, hence the vectors are necessarily independent. So set

$$W = \begin{bmatrix} W_A & 0 \\ 0 & W_B \end{bmatrix}.$$

The reader may verify that

$$W^{-1} = \begin{bmatrix} \frac{-1+\sqrt{3}\,i}{6} & \frac{-1-\sqrt{3}\,i}{6} & 1/3 & 0 & 0 \\ \frac{-1-\sqrt{3}\,i}{6} & \frac{-1+\sqrt{3}\,i}{6} & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

Letting

$$X = e \, \frac{-3 - \sqrt{3}\, i}{2} \, t , \qquad Y = e \, \frac{-3 + \sqrt{3}\, i}{2} \, t \ ,$$

it follows that

$$P(t; f^{\infty}) = \begin{bmatrix} \frac{1}{3} + \frac{1}{3} X + \frac{1}{3} Y & \frac{1}{3} + \frac{-1+\sqrt{3}\, i}{6} X + \frac{-1-\sqrt{3}\, i}{6} Y & \frac{1}{3} + \frac{-1-\sqrt{3}\, i}{6} X + \frac{-1+\sqrt{3}\, i}{6} Y & 0 & 0 \\ \frac{1}{3} + \frac{-1-\sqrt{3}\, i}{6} X + \frac{-1+\sqrt{3}\, i}{6} Y & \frac{1}{3} + \frac{1}{3} X + \frac{1}{3} Y & \frac{1}{3} + \frac{-1+\sqrt{3}\, i}{6} X + \frac{-1-\sqrt{3}\, i}{6} Y & 0 & 0 \\ \frac{1}{3} + \frac{-1+\sqrt{3}\, i}{6} X + \frac{-1-\sqrt{3}\, i}{6} Y & \frac{1}{3} + \frac{-1-\sqrt{3}\, i}{6} X + \frac{-1+\sqrt{3}\, i}{6} Y & \frac{1}{3} + \frac{1}{3} X + \frac{1}{3} Y & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 - e^{-t} & e^{-t} \end{bmatrix} ,$$

and that

$$P(t;f^{\infty})r(f) = \begin{bmatrix} 4 + X + Y \\ 4 + \frac{-1 - \sqrt{3}\,i}{2} X + \frac{-1 + \sqrt{3}\,i}{2} Y \\ 4 + \frac{-1 + \sqrt{3}\,i}{2} X + \frac{-1 - \sqrt{3}\,i}{2} Y \\ 4 \\ 4 \end{bmatrix} .$$

Now,

$$\begin{aligned} P_1(t;f^{\infty})r(f) &= 4 + X + Y \\ &= 4 + e^{-3/2t}[e^{-\sqrt{3}/2\ it} + e^{\sqrt{3}/2\ it}] \\ &= 4 + 2e^{-3/2t} \cos \frac{\sqrt{3}}{2} t . \end{aligned}$$

Hence,

$$V_1^t(f^{\infty}) = \int_0^t [4 + 2e^{-3/2s} \cos \frac{\sqrt{3}}{2} s]ds$$

and integrating by parts (twice), gives

$$V_1^t(f^{\infty}) = 4t + e^{-3/2t} \frac{\sqrt{3}}{3} \sin \frac{\sqrt{3}}{2} t - e^{-3/2t} \cos \frac{\sqrt{3}}{2} t + 1 .$$

Similarly for the remaining components, it follows that

$$V^t(f^\infty) = t\begin{bmatrix}4\\4\\4\\4\\4\end{bmatrix} + e^{-3/2t}\sin\frac{\sqrt{3}}{2}t\begin{bmatrix}\sqrt{3}/3\\\sqrt{3}/3\\-2\sqrt{3}/3\\0\\0\end{bmatrix}$$

$$+ e^{-3/2t}\cos\frac{\sqrt{3}}{2}t\begin{bmatrix}-1\\1\\0\\0\\0\end{bmatrix} + \begin{bmatrix}1\\-1\\0\\0\\0\end{bmatrix}$$

for all $t \geq 0$.

It then follows that

$$\delta(t) = e^{-3/2t}[\cos\frac{\sqrt{3}}{2}t - \frac{\sqrt{3}}{3}\sin\frac{\sqrt{3}}{2}t],$$

which changes sign at the set of times

$$\{t \geq 0: t = 2\pi\sqrt{3}/9 + n(2\pi\sqrt{3}/3),\ n = 0, 1, 2, \ldots\},$$

and is non-zero for all other times $t \geq 0$. So $\delta(t)$ has the oscillation property referred to above and the example is complete.

## 4. Existence of Stationary Optimal Policies

The existence of stationary optimal policies on recurrent states will be characterized in terms of the vector $V$. For discrete time processes Lanery [10] has studied optimal policies in relation to the terminal reward vector $v$.

For $x, z \in R^N$ and $M$ a subset of $\{1, 2, \ldots, N\}$, say $x = z$ $(x \leq z)$ on $M$ if $x_i = z_i$ $(x_i \leq z_i)$ for all $i \in M$. Also $x < z$ on $M$ if $x \leq z$ on $M$ and $x_i < z_i$ for at least one $i \in M$. A decision rule $f$ equals a decision rule $g$ on the set $M$ if $f(i) = g(i)$ for all $i \in M$.

Theorem 4. If $V = v^*$, then for any $f \in D(v)$, $V^t(f^\infty, v) = V^t(\pi^*, v)$ on $C(f)$ for all $t \geq 0$.

Proof: Let $f \in D(v)$ and let $\pi^*$ be any optimal policy. Define

$$t_1 = \begin{cases} \infty, & \text{if } V^t(f^\infty, v) = V^t(\pi^*, v) \text{ on } C(f) \text{ for all } t \geq 0 \\ \inf\{t \geq 0 : V^t(\pi^*, v) > V^t(f^\infty, v) \text{ on } C(f)\}, & \text{otherwise} \end{cases} .$$

Assume $t_1 < \infty$. It will be established that this gives a contradiction hence implying that $t_1 = \infty$.

Let

(4) $$x_1 = V^{t_1}(\pi^*, v) > V^{t_1}(f^\infty, v) = x_2 \quad \text{on} \quad C(f).$$

Since $f \in F'$, from Lemma 1, (iii), I,

$$V^t(f^\infty, x_2) - tU \to y(f) + P^*(f)x_2 \quad \text{as} \quad t \to \infty,$$

and

$$V^t(f^\infty, v) - tU \to y(f) + P^*(f)v \quad \text{as} \quad t \to \infty.$$

Now

$$V^t(f^\infty, x_2) = V^{t+t_1}(f^\infty, v) \quad \text{for all} \quad t \geq 0,$$

so

$$y(f) + P^*(f)v = y(f) + P^*(f)x_2 - t_1U \cdot$$

Also since

$$f \in D(v) \quad \text{and} \quad v^* = V, \quad y(f) + P^*(f)v = V,$$

so

(5) $$V = y(f) + P^*(f)x_2 - t_1 \cdot$$

From (4) and the structure of $P^*(\cdot)$ it follows that $P^*(f)x_1 > P^*(f)x_2$, so that (5) implies

(6) $$y(f) + P^*(f)x_1 - t_1U > V \cdot$$

Let

$$\pi' = \begin{cases} \pi^*, & 0 \leq t < t_1 \\ f, & t_1 \leq t \end{cases} .$$

Then $V^t(\pi^*, v) - V^t(\pi', v) \rightarrow V - (y(f)+P^*(f)x_1 - t_1U)$ as $t \rightarrow \infty$. Since $V^t(\pi^*, v) \geq V^t(\pi', v)$ for all $t \geq 0$, it follows that $V - (y(f)+P^*(f)x_1 - t_1U) \geq 0$, which contradicts (6). Conclude that $t_1 = \infty$, and the theorem follows.

The next result may be viewed as the "necessary condition" counterpart of Theorem 4.

Theorem 5. If for some $f \in F$, $V^t(f^\infty, v) = V^t(\pi^*, v)$ on $C(f)$ for all $t \geq 0$, then $V = v^*$ on $C(f)$.

Proof: Since

$$\lim_{t \rightarrow \infty} \frac{1}{t} V^t(f^\infty, v) = P^*(f)r(f)$$

and

$$\lim_{t \rightarrow \infty} \frac{1}{t} V^t(\pi^*, v) = U ,$$

it follows that $P^*(f)r(f) = U$ on $C(f)$. Hence by Theorem 2, I, and Lemma 1, (iii), I,

$$V = y(f) + P^*(f)v \quad \text{on} \quad C(f).$$

The structure of $Q(f)$ on $C(f)$ together with (4), I implies that $r(f) + Q(f)V = U$ on $C(f)$ and $Q(f)U = 0$ on $C(f)$. Now let $h' \in F^*$ and define

$$h(i) = \begin{cases} f(i), & i \in C(f) \\ h'(i), & \text{otherwise} \end{cases} .$$

Then $h \in F^*$ and consequently $h \in F'$. Now from (3), I, and (4), I, and the structure of $P^*(\cdot)$ on $C(f)$ it follows that $y(f) = y(h)$ on $C(f)$ and $P^*(h)v = P^*(f)v$ on $C(f)$. Hence

$$(7) \qquad V = y(f) + P^*(f)v = y(h) + P^*(h)v \quad \text{on} \quad C(f).$$

Now from Lemma 1, (iii), I,

$$(8) \qquad V^t(h^\infty, v) - tU \to y(h) + P^*(h)v \quad \text{as} \quad t \to \infty.$$

Hence from (7) and (8), and since $h \in F'$

$$V \geq \sup_{(\cdot)F'} (y(\cdot)+P^*(\cdot)v) = v^* \geq y(h) + P^*(h)v .$$

So by (7), $V = v^*$ on $C(f)$.

A decision rule $f$ is said to be recurrent if $C(f) = S$, i.e., if the Markov process associated with the policy $f^\infty$ has no transient states. In general a recurrent decision rule may have several communicating classes. Recurrent processes are assumed by many authors in elementary textbooks on standard Markov chain theory and represent a large class of problems.

With the assumption of no transient states more powerful results are usually obtained. In some cases one need only assume that certain sets of decision rules are recurrent. For the two preceding theorems, an assumption of this nature yields the result stated next.

Theorem 6. Assume $f$ is recurrent for all $f$ in $D(v)$. Then the following are equivalent.

(i) There exists a stationary optimal policy.

(ii) For any $f \in D(v)$, $f^{\infty}$ is optimal.

(iii) $V = v^*$.

Proof: By Theorem 4, condition (iii) implies (ii). Also, (ii) certainly implies (i).

Let $g^{\infty}$ be a stationary optimal policy. Then

$$V^{t}(g^{\infty}, v) - tU \longrightarrow y(g) + P^*(g)v = V \quad \text{as} \quad t \longrightarrow \infty.$$

It follows that $V = v^*$ and (i) implies (iii).

Remark 1. If the rewards earned using $\pi^*$ up to some time point $\bar{t}$ are known, say $x_{\bar{t}} = V^{\bar{t}}(\pi^*, v)$, then by setting $v = x_{\bar{t}}$, the preceding theorems provide necessary and sufficient conditions for achieving maximal expected rewards on various recurrent states by using initially stationary policies.

Corollary 3. If $f \notin F^* \cap D(v)$, then $f^\infty$ is not an optimal policy.

Proof: Assume $f^\infty$ is optimal. Then since $f \in F'$, $y(f) + P^*(f)v = V \geq v^* \geq y(f) + P^*(f)v$, which implies $f \in D(v)$. Further, it is clear that $f \in F^*$. Hence $f \in F^* \cap D(v)$, a contradiction.

Remark 2. In view of the preceding corollary, it follows that if $F^* \cap D(v) = \phi$ then there is no stationary optimal policy. For an example where $F^* \cap D(v) = \phi$, the reader may see the three state example of Lembersky in [13]. A further implication of this corollary is that if $F^* \cap D(V^t(\pi^*, v)) = \phi$ for all $t \geq 0$, then there does not exist an initially stationary optimal policy.

As noted in Example 3, $F^* = \{f, g\}$ while $\pi^*(t) = f$ for all $t \geq 0$. Also, consideration of $y(f)$ and $y(g)$ reveals that $D(v) = D(0) = \{f\}$. Thus $F^* \cap D(v) = \{f\}$. Also recall in observation 3, following Example 3, that if the algorithm for computing a decision rule $h$ as in Lemma 3, (ii), starts with $g$, it gives $g \in F^*$ and fails to give $f$. In light of Corollary 3, it would appear that when there exists a stationary optimal policy, the set $F^* \cap D(v)$ represents a desirable refinement of the set $F^*$.

The next corollary follows immediately from Corollary 3 and Theorem 4 and relates to statements of the preceding paragraph.

Corollary 4. Assume $F^* \cap D(v) = \{f\}$ and $V = v^*$. Then $V^t(f^\infty, v) = V^t(\pi^*, v)$ on $C(f)$ for all $t \geq 0$. Further, if $f$ is recurrent, then $f^\infty$ is the unique stationary optimal policy.

Recall that in Counterexample 4, $\pi^*(t)$ was constant (stationary) on $C$ but continued to switch off of $C$. Using Corollary 5 it is possible in such situations to obtain a decision rule that is in $F^*$.

Corollary 5. Assume for some $g \in F$ that $V^t(g^\infty, v) = V^t(\pi^*, v)$ on $C$ for all $t \geq 0$ and that $C(g) \subset C$. If $f \in D(y(g)+P^*(g)v)$, then $f \in F^*$.

Proof: From Theorem 5, $V = v^*$ on $C(g)$. Let $f \in D(y(g)+P^*(g)v)$. From Theorem 2, I, $V = V^*$, hence it follows that $V = \sup_{(\cdot)F'} (y(\cdot)+P^*(\cdot)V)$. Since $P^*(g)r(g) = U$ on $C$, from Lemma 1, (iii), I, $y(g) + P^*(g)v = V$ on $C$. Also for each $(\cdot) \in F'$, $C(\cdot) \subset C$, so it follows that

$$\begin{aligned} V &= \sup_{(\cdot)\in F'} (y(\cdot)+P^*(\cdot)(y(g)+P^*(g)v)) \\ &= y(f) + P^*(f)(y(g)+P^*(g)v) \ . \end{aligned}$$

Since $f \in F'$, $C(f) \subset C$, hence

$$y(f) + P^*(f)(y(g)+P^*(g)v) = y(f) + P^*(f)V \ .$$

Then from Theorem 4, I, $f \in F^*$.

## 5. The $N = 3$ Case

As previously noted in Section 3, when $N = 2$, there exists an initially stationary optimal policy. Observe that when $N = 2$, since zero is necessarily an eigenvalue for $Q(\cdot)$ and since complex eigenvalues occur in conjugate pairs it follows that for each $f \in F$, $Q(f)$ has real eigenvalues. Also, in Example 4, some of the eigenvalues for the matrices $Q(f)$ and $Q(g)$ were complex. The effect of these complex eigenvalues was that $\delta(t)$, $t \geq 0$ was a function involving sines and cosines and continued to oscillate above and below zero as a function of $t$ and hence there was no initially stationary optimal policy. Thus it would appear that to guarantee the existence of an initially stationary optimal policy, one must rule out the possibility of complex eigenvalues for the matrices $Q(\cdot)$.

Several assumptions can be made in order to have only real eigenvalues. From Theorem 5,I a possible assumption is that for all decision rules used beyond $t^*$ in some $\pi^*$, the corresponding $Q(\cdot)$ matrices have real eigenvalues. Other sets of decision rules could also be assumed to have $Q(\cdot)$ matrices of real eigenvalues, namely $F$, $\{f \in F: Q(f)U = 0\}$, $F'$, or $F^*$. Using the assumption that for $f \in F^*$ the eigenvalues for $Q(f)$ are real appears to have potential value as part of the suitable hypothesis that will guarantee the existence of an initially stationary optimal policy. While unable to develop a definitive answer, the remainder of this section illustrates a potential approach for solving this problem. Specifically, to indicate the type of arguments involved, it will be established for $N = 3$ that if $Q(f)$ has real eigenvalues for all $f \epsilon F^*$, then the optimal policy may not oscillate between two decision rules as a function of time.

<u>Remark 3.</u> For $N = 3$, assume for $f \epsilon F^*$ that $Q(f)$ has real eigenvalues. Assert for the optimal policy $\pi^*$ and any $t_0 \geq t^*$ (where $t^*$ is selected as in Theorem 5, I), if

$$\pi^*(t) = \begin{cases} f & t_0 \leq t < t_1 < \infty \\ g & t_1 \leq t < t_2 < \infty \\ h & t_2 \leq t < t_3 \leq \infty \end{cases},$$

then $f \neq h$. In other words, beyond $t^*$ an optimal policy cannot oscillate between two decision rules (as happened in Example 4). This fact does not rule out the possibility that $\pi^*(t)$ may equal $f$ on some interval beyond $t_3$, hence it does not guarantee an initially stationary optimal policy exists.

To establish Remark 3 the notation below and the lemmas which follow (and are true for any $N$) will be useful.

For $f, g \epsilon F$ and $x \epsilon R^N$, let

$$\phi^t(g,f;x) = r(g) + Q(g)V^t(g^\infty, x) - r(f) - Q(f)V^t(g^\infty, x) \tag{9}$$

for all $t \geq 0$.

Lemma 4. Assume $f, g \in F^*$. If $g$ is single class, then for any $x \in R^N$,

$$\phi^t(g,f;x) \rightarrow 0 \quad \text{as} \quad t \rightarrow \infty.$$

Proof: Since $Q(g)P^*(g) = 0$ and $g$ is single class, it follows from Lemma 1, (iii), I, as $t \rightarrow \infty$, that

$$\phi^t(g,f;x) \rightarrow r(g) + Q(g)[y(g)+P^*(g)x] - r(f)$$
$$- Q(f)[y(g)+P^*(g)x] = U - r(f) - Q(f)y(g).$$

From Theorem 4, I, $V = y(g) + k \circ 1$ for some scalar $k$. Hence

$$U - r(f) - Q(f)y(g) = U - r(f) - Q(f)V$$

and since $f \in F^*$, from Theorem 4, I, $U - r(f) - Q(f)y(g) = 0$. Hence $\phi^t(g,f;x) \rightarrow 0$ as $t \rightarrow \infty$.

The next lemma is stated for stationary policies. A more general version for any measurable policies is given by Miller [14] in his proof of Theorem 1, I for the opposite time orientation and with $x = 0$. The argument below is essentially the same as Miller's, only now $x$ is allowed to be nonzero.

Lemma 5. Let $f$ and $g \in F$ and $x \in R^N$. Then for any $t \geq 0$,

$$V^t(g^\infty, x) - V^t(f^\infty, x) = \int_0^t P(s;f^\infty)\phi^{t-s}(g,f;x)ds .$$

Proof: Fix $t \geq 0$ and define

$$A(s) = [P(s;g^\infty)-P(s;f^\infty)]V^{t-s}(g^\infty, x) \quad \text{for all} \quad 0 \leq s \leq t.$$

Since $P(0;\cdot) = I$ and $V^0(\cdot, x) = x$, $A(0) = 0$ and $A(t) = [P(t;g^\infty)-P(t;f^\infty)]x$. Now $P(s;\cdot)$ and $V^{t-s}(\cdot;x)$ are absolutely continuous in $s$ (see [14]), so $A(s)$ is absolutely continuous in $s$ and must then equal the integral of its derivative. So

$$0 = \int_0^t \frac{d}{ds}A(s)ds - A(t).$$

From (1), I, $\frac{d}{ds}P(s;\cdot) = P(s;\cdot)Q(\cdot)$ and from Lemma 1, (ii), I,

$$\frac{d}{ds} V^{t-s}(g^\infty, x) = -[r(g)+Q(g)V^{t-s}(g^\infty, x)] .$$

So

$$(10) \qquad 0 = \int_0^t \left\{[P(s;f^\infty)-P(s;g^\infty)]r(g) + P(s;f^\infty)[Q(g)-Q(f)] \times V^{t-s}(g^\infty, x)\right\} ds - A(t).$$

From (2), I,

$$V^t(g^\infty, x) - V^t(f^\infty, x) = \int_0^t [P(s;g^\infty)r(g) - P(s;f^\infty)r(f)]ds + [P(t;g^\infty) - P(t;f^\infty)]x .$$

Using (10) to substitute for $\int_0^t P(s;g^\infty)r(g)ds$, the lemma follows.

Proof of Remark 3: Let $\pi^*$ be any optimal policy and let $f, g, h$ and $t_0, t_1, t_2$ be as in the remark. Then by Theorem 5, I, $f, g, h \in F^*$. By standard results in matrix theory, zero is an eigenvalue for $Q(g)$ and the remaining eigenvalues are less than or equal to zero. Also, the number of non-zero eigenvalues must be less than or equal to two. Let $-\lambda_1 \leq -\lambda_2 \leq 0$ denote the possibly non-zero eigenvalues for $Q(g)$.

Let $x = V^{t_1}(\pi^*, v)$. Assert that $\phi^t(g, f;x) > 0$ for all $t > 0$.

Note that if $f(i) = g(i)$, then $\phi_i^t(g, f;x) = 0$ for all $t \geq 0$. Also, there exists an $i^*$ such that $f(i^*) \neq g(i^*)$ and such that $\phi_{i^*}^t(g, f;x) > 0$ for all $t > 0$ sufficiently small. This follows since the definition of $x$ and Theorem 1, (ii), I imply $\phi^t(g, f;x) \geq 0$ for all $0 \leq t < t_2 - t_1$, since $\phi^t(\cdot, \cdot;x)$ is continuous in $t$, and since Lemma 5 implies that if $\phi^t(g, f;x) = 0$ for all $t > 0$ sufficiently small, then $V^t(g^\infty, x) = V^t(f^\infty, x)$ for all $t > 0$ sufficiently small, which would contradict the need to switch from $f$ to $g$ at $t_1$. For simplicity of exposition in establishing the assertion, assume without loss of generality that $i^* = 1$. To establish the assertion the following cases involving $\lambda_1$ and $\lambda_2$ will be considered.

Case 1. Assume $\lambda_2 = 0$.

From the Appendix, (3) is follows that $\phi_1^t(g,f;x) = k + a_1 e^{-\lambda_1 t}$ for all $t \geq 0$.

From the definition of $t_1$ and Theorem 1, I, $\phi_1^0(g,f;x) = 0$, which implies that $a_1 = -k$. Hence $\phi_1^t(g,f;x) = k - ke^{-\lambda_1 t}$ for all $t \geq 0$. Therefore either

$$\phi_1^t(g,f;x) \begin{cases} > 0 & \text{for all} \quad t > 0 \\ = 0 & \text{for all} \quad t \geq 0 \\ < 0 & \text{for all} \quad t > 0 \end{cases}$$

But since $\phi_1^t(g,f;x) > 0$ for all $t > 0$ sufficiently small, it follows that $\phi_1^t(g,f;x) > 0$ for all $t > 0$.

Case 2. Assume $-\lambda_1 = -\lambda_2 < 0$.

From the Appendix, (4), $\phi_1^t(g,f;x) = k + a_1 te^{-\lambda_1 t} + a_2 e^{-\lambda_1 t}$.

Since $-\lambda_2 < 0$ and since zero is an eigenvalue for each recurrent class, $g$ must then be single class. Hence from Lemma 4,

$$\phi_1^t(g,f;x) \rightarrow 0 \quad \text{as} \quad t \rightarrow \infty$$

which implies that $k = 0$. Also since

$$\phi_1^0(g,f;x) = 0, \quad a_2 = 0, \quad \text{so} \quad \phi_1^t(g,f;x) = a_1 te^{-\lambda_1 t}.$$

Again either

$$\phi_1^t(g,f;x) \begin{cases} > 0 & \text{for all} \quad t > 0 \\ = 0 & \text{for all} \quad t \geq 0 \\ < 0 & \text{for all} \quad t > 0 \end{cases},$$

and arguing as above, it follows that

$$\phi_1^t(g,f;x) > 0 \quad \text{for all} \quad t > 0.$$

Case 3. Assume $-\lambda_1 < -\lambda_2 < 0$.

The nonzero eigenvalues for $Q(g)$ distinct imply $g$ is single class and from Appendix, (2),

$$\phi_1^t(g,f;x) = k + a_1 e^{-\lambda_1 t} + a_2 e^{-\lambda_2 t}.$$

From Lemma 4, $\phi_1^t(g,f;x) \to 0$ as $t \to \infty$, so $k = 0$. Also, $\phi_1^t(g,f;x) = 0$ for $t = 0$ so $a_2 = -a_1$. Hence

$$\phi_1^t(g,f;x) = a_1 e^{-\lambda_1 t} - a_1 e^{-\lambda_2 t}.$$

Since $e^{-\lambda_1 t} - e^{-\lambda_2 t} < 0$ for all $t > 0$, again either

$$\phi_1^t(g,f;x) \begin{cases} > 0 & \text{for all} \quad t > 0 \\ = 0 & \text{for all} \quad t \geq 0 \\ < 0 & \text{for all} \quad t > 0 \end{cases},$$

and so $\phi_1^t(g, f; x) > 0$ for all $t > 0$.

Thus $\phi_1^t(g, f; x) > 0$ for all $t > 0$ establishing the assertion that

$$\phi^t(g, f; x) > 0 \quad \text{for all} \quad t > 0.$$

From Theorem 1, I, it follows that since $\pi^*$ next switches from $g$ to $h$, $\phi^{t_2 - t_1}(g, h; x) = 0$. Therefore $f \neq h$, completing the proof of Remark 3.

# BIBLIOGRAPHY

[1] Bartholomew, D.J. (1969), "Sufficient Conditions for a Mixture of Exponentials to be a Probability Density Function," *Ann. Math. Stat.* 40, 2183-2188.

[2] Becker, N.G. (1970), "Control of a Pest Population," *Biometrics*, 26, 365-375.

[3] Chiang, C.L. (1968), *Introduction to Stochastic Processes in Biostatistics*, Wiley, New York.

[4] Doob, J. (1953), *Stochastic Processes*, Wiley, New York.

[5] Franklin, J.N. (1968), *Matrix Theory*, Prentice Hall, Englewood Cliffs, NJ.

[6] Gantmacher, F.R. (1959), *Applications of the Theory of Matrices*, Interscience Publishers Inc., New York.

[7] Jaquette, D.L. (1972), "Mathematical Models for Controlling Growing Biological Populations: A Survey," *Operations* Research, 20, 1142-1151.

[8] Karlin, S. (1969), *A First Course in Stochastic Processes*, Academic Press, New York.

[9] Lanery, E. (1967), "Etude Asymptotique des Systemes Markoviens a Commande," R.I.R.O. 3, 3-56.

[10] Lanery, E. (1968), "Complements a l'Etude Asymptotique des Systemes Markoviens a Commande," Institut de Recherche d' Informatique et d'automatique, Rocquencourt, France.

[11] Lembersky, M.R. (1972), "Initially Stationary $\epsilon$-Optimal Policies in Continuous Time Markov Decision Chains," Technical Report No. 22, Department of Operations Research, Stanford University, Stanford, CA.

[12] Lembersky, M.R. (1974), "On Maximal Rewards and $\epsilon$-Optimal Policies in Continuous Time Markov Decision Chains," *Ann. Statist.*, 2, 159-169.

[13] Lembersky, M.R. (1974), "Preferred Rules in Continuous Time Markov Decision Processes," Management Science, (In Print).

[14] Miller, B.L. (1968), "Finite State Continuous Time Markov Decision Processes with a Finite Planning Horizon," Siam J. Control 6, 266-280.

[15] Miller, B.L. (1968), "Finite State Continuous Time Markov Decision Processes with an Infinite Planning Horizon," J. Math. Anal. Appl. 22, 552-569.

[16] Teghem, J., Jr. (1971), "Processus de Decision Markovien: Tactiques Bias-Optimal," CCERO (Belgium) 13, 124-140.

[17] Veinott, A.F., Jr. (1966), "On Finding Optimal Policies in Discrete Dynamic Programming with No Discounting," Ann. Math. Stat. 37, 1284-1294.

[18] Veinott, A.F., Jr. (1969), "Discrete Dynamic Programming with Sensitive Discount Optimality Criteria," Ann. Math. Stat. 40, 1635-1660.

# APPENDIX

# APPENDIX

To describe the form of $\phi^t(g,f;x)$ needed in Section 5, II, when $Q(g)$ is assumed to have real eigenvalues, it is convenient to find $P(t;g^\infty)$ first.

Chiang [3] assumes no absorbing states and that the eigenvalues for $Q(g)$ are real and distinct. Under these assumptions he gives an explicit solution for $P(t;g^\infty)$. Assume $\rho_0 = 0, \rho_2, \ldots, \rho_{N-1} < 0$ and distinct. Define for $k = 0, 1, \ldots, N-1$

$$A'(k) = \begin{bmatrix} \rho_k - q_{11} & -q_{21} & \cdots & -q_{N1} \\ -q_{12} & \rho_k - q_{22} & \cdots & -q_{N2} \\ \vdots & \vdots & & \\ -q_{1N} & -q_{2N} & \cdots & \rho_k - q_{NN} \end{bmatrix},$$

where $q_{ij} = Q_{ij}(g)$. Then

$$P_{ij}(t;g^\infty) = \sum_{k=0}^{N-1} \frac{A'_{ij}(k)e^{\rho_k t}}{\prod_{\substack{m=0 \\ m \neq k}}^{N-1} (\rho_k - \rho_m)}, \quad i,j = 1, \ldots, N.$$

Also by letting $A_{ij}(\ell)$ be the cofactor of the matrix $A(\ell) = [\rho_\ell I - Q(g)]$ and

$$W_\ell(k) = \begin{bmatrix} A_{k1}(\ell) \\ \vdots \\ A_{kN}(\ell) \end{bmatrix}$$

by an eigenvector of $Q(g)$ for $\rho = \rho_\ell$, if $W(k) = (W_1(k), \ldots, W_N(k))$, then $P(t;g^\infty) = W(k)\Lambda(t)W^{-1}(k)$, where $\Lambda(t)$ is a diagonal matrix with entries $1, e^{\rho_2 t}, \ldots, e^{\rho_{N-1} t}$. Chiang then gives

$$(1) \qquad P_{ij}(t;g^\infty) = \sum_{\ell=0}^{N-1} A_{ki}(\ell) \frac{W_{j\ell}(k)}{|W(k)|} e^{\rho_\ell t}, \quad i,j = 1, \ldots, N.$$

To simplify notation, write $W(k) = W$, then the form $P(t;g^\infty) = W\Lambda(t)W^{-1}$ may always be used when $Q(g)$ has distinct eigenvalues [5]. If zero is a multiple eigenvalue and the remaining nonzero eigenvalues are distinct, by rearranging the rows of $Q(g)$ and partitioning according to the recurrent classes and transient states, i.e.,

$$Q(g) = \begin{bmatrix} Q_1 & & 0 & \\ & Q_2 & & 0 \\ 0 & & \ddots\, Q_k & \\ & Q_{TR} & & Q_T \end{bmatrix},$$

then

$$P(t;g^\infty) = \begin{bmatrix} P_1(t) & & 0 & \\ & P_2(t) & & 0 \\ 0 & & \ddots\, P_k(t) & \\ P_{TR}(t) & & & P_T(t) \end{bmatrix}$$

and the form $P(t;g^{\infty}) = W\Lambda(t)W^{-1}$ may still be used. Note that this procedure was used in Counterexample 4, II.

If the nonzero eigenvalues are not distinct it is still possible to find $P(t;g^{\infty})$ by using Jordan's Theorem stated next. This statement of Jordan's Theorem is taken from Franklin [5].

<u>Jordan's Theorem.</u> Let $Q(g)$ be $N \times N$ with eigenvalues $\lambda_1, \ldots, \lambda_s$ with multiplicities $m_1, \ldots, m_s$,

$$\det(\lambda I - Q(g)) = \prod_{j=1}^{s} (\lambda - \lambda_j)^{m_j} .$$

Then $Q(g)$ is similar to a matrix of the form

$$J = \begin{bmatrix} \Lambda_1 & & \\ & \Lambda_2 & 0 \\ 0 & & \ddots \; \Lambda_s \end{bmatrix},$$

where $\Lambda_i$ is $m_i \times m_i$ and is of the form

$$\Lambda_i = \begin{bmatrix} \lambda_i & & & 0 \\ * & \lambda_i & & \\ & & \ddots \; \lambda_i & \\ 0 & & * & \lambda_i \end{bmatrix}$$

and each $*$ equals zero or one.

Franklin then gives that $P_{ij}(t;f^{\infty})$ is of the form

$$\sum_{k=1}^{s} \sum_{\ell=0}^{m_k-1} \xi_{ijk\ell} t^{\ell} \quad e^{\lambda_k t}$$

for some scalars $\xi_{ijk\ell}$ .

Assume now that for $N = 3$, $Q(g)$ has the distinct real eigenvalues, $0$, $-\lambda_1 < -\lambda_2 < 0$. Using the form $P(t;g^{\infty}) = W\Lambda(t)W^{-1}$, from (1)

$$P(t;f^{\infty}) = \begin{bmatrix} k_{11} + \sum_{i=1}^{2} a_{i1} e^{-\lambda_i t} & k_{12} + \sum_{i=1}^{2} a_{i2} e^{-\lambda_i t} & k_{13} + \sum_{i=1}^{2} a_{i3} e^{-\lambda_i t} \\ k_{21} + \sum_{i=1}^{2} b_{i1} e^{-\lambda_i t} & k_{22} + \sum_{i=1}^{2} b_{i2} e^{-\lambda_i t} & k_{23} + \sum_{i=1}^{2} b_{i3} e^{-\lambda_i t} \\ k_{31} + \sum_{i=1}^{2} c_{i1} e^{-\lambda_i t} & k_{32} + \sum_{i=1}^{2} c_{i2} e^{-\lambda_i t} & k_{33} + \sum_{i=1}^{2} c_{i3} e^{-\lambda_i t} \end{bmatrix}$$

with the obvious designations for $k_{ij}$, $a_{ij}$, $b_{ij}$, $c_{ij}$, $i = 1,2$, $j = 1,2,3$. Then

$$\int_0^t P(s;g^{\infty})r(g)ds = \begin{bmatrix} \int_0^t \sum_{j=1}^{3} r_j(g)\left[k_{1j} + \sum_{i=1}^{2} a_{ij}e^{-\lambda_i s}\right]ds \\ \int_0^t \sum_{j=1}^{3} r_j(g)\left[k_{2j} + \sum_{i=1}^{2} b_{ij}e^{-\lambda_i s}\right]ds \\ \int_0^t \sum_{j=1}^{3} r_j(g)\left[k_{3j} + \sum_{i=1}^{2} c_{ij}e^{-\lambda_i s}\right]ds \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^{3} r_j(g)\left[k_{1j}t - \sum_{i=1}^{2}\left[\frac{a_{ij}}{\lambda_i}e^{-\lambda_i t} + \frac{a_{ij}}{\lambda_i}\right]\right] \\ \sum_{j=1}^{3} r_j(g)\left[k_{2j}t - \sum_{i=1}^{2}\left[\frac{b_{ij}}{\lambda_i}e^{-\lambda_i t} + \frac{b_{ij}}{\lambda_i}\right]\right] \\ \sum_{j=1}^{3} r_j(g)\left[k_{3j}t - \sum_{i=1}^{2}\left[\frac{c_{ij}}{\lambda_i}e^{-\lambda_i t} + \frac{c_{ij}}{\lambda_i}\right]\right] \end{bmatrix} .$$

Also for $x \in R^3$ ,

$$P(t;g^{\infty})x = \begin{bmatrix} \sum_{j=1}^{3} x_j\left[k_{1j} + \sum_{i=1}^{2} a_{ij}e^{-\lambda_i t}\right] \\ \sum_{j=1}^{3} x_j\left[k_{2j} + \sum_{i=1}^{2} b_{ij}e^{-\lambda_i t}\right] \\ \sum_{j=1}^{3} x_j\left[k_{3j} + \sum_{i=1}^{2} c_{ij}e^{-\lambda_i t}\right] \end{bmatrix} .$$

Assume that $f(1) \neq g(1)$ and that $[Q(g)-Q(f)]_1 = (q_1\ q_2\ q_3)$. Then

$$\phi_1^t(g,f;x) = r_1(g) - r_1(f) + (q_1\ q_2\ q_3)\left[\int_0^t P(s;g^{\infty})r(g)ds + P(t;g^{\infty})x\right] .$$

By multiplying and collecting terms it follows that $\phi_1^t(g,f;x)$ may be written in the form, for some $k$, $a_1$, and $a_2 \in R^1$,

(2) $$\phi_1^t(g,f;x) = k + a_1e^{-\lambda_1 t} + a_2e^{-\lambda_2 t} \quad \text{for all} \quad t \geq 0$$

where $-\lambda_1 < -\lambda_2 < 0$ are the distinct nonzero real eigenvalues of $Q(g)$.

Assume next that for $N = 3$, $Q(g)$ has zero as a multiple eigenvalue, i.e., $-\lambda_1 \leq -\lambda_2 = 0$. The preceding argument for (2) may be used with $\lambda_2 = 0$ to give $\phi_1^t(g,f;x)$ in the form, for some $k$ and $a_1$ in $R^1$,

$$\text{(3)} \qquad \phi_1^t(g,f;x) = k + a_1 e^{-\lambda_1 t} \quad \text{for all} \quad t \geq 0$$

where $-\lambda_1 \leq -\lambda_2 = 0$ and $-\lambda_1$, 0, 0 are the real eigenvalues for $Q(g)$.

Finally for $N = 3$ assume that $Q(g)$ has $-\lambda_1 = -\lambda_2 < 0$ as real nonzero eigenvalues. From Jordan's Theorem $P(t; g^\infty)$ is now of the form

$$\begin{bmatrix} k_{11}+a_{11}te^{-\lambda_1 t}+a_{21}e^{-\lambda_1 t} & k_{12}+a_{12}te^{-\lambda_1 t}+a_{22}e^{-\lambda_1 t} & k_{13}+a_{13}te^{-\lambda_1 t}+a_{23}e^{-\lambda_1 t} \\ k_{21}+b_{11}te^{-\lambda_1 t}+b_{21}e^{-\lambda_1 t} & k_{22}+b_{12}te^{-\lambda_1 t}+b_{22}e^{-\lambda_1 t} & k_{23}+b_{13}te^{-\lambda_1 t}+b_{23}e^{-\lambda_1 t} \\ k_{31}+c_{11}te^{-\lambda_1 t}+c_{21}e^{-\lambda_1 t} & k_{32}+c_{12}te^{-\lambda_1 t}+c_{22}e^{-\lambda_1 t} & k_{33}+c_{13}te^{-\lambda_1 t}+c_{33}e^{-\lambda_1 t} \end{bmatrix}$$

with new designations for $k_{ij}$, $a_{ij}$, $b_{ij}$ and $c_{ij}$. Following the same procedures used when $-\lambda_1 < -\lambda_2 < 0$ it follows that $\phi_1^t(g,f;x)$ may be written as

$$\text{(4)} \qquad \phi_1^t(g,f;x) = k + a_1 te^{-\lambda_1 t} + a_2 e^{-\lambda_1 t} \quad \text{for all} \quad t \geq 0$$

where $-\lambda_1 = -\lambda_2 < 0$ are the real nonzero eigenvalues for $Q(g)$ and $k$, $a_1$ and $a_2$ are in $R^1$.