stichting
mathematisch
centrum

MC

AFDELING BESLISKUNDE BW 71/77 MAART
(DEPARTMENT OF OPERATIONS RESEARCH)

P.J. SCHWEITZER & A. FEDERGRUEN

THE FUNCTIONAL EQUATIONS OF UNDISCOUNTED MARKOV RENEWAL PROGRAMMING

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): 90C45

The Functional Equations of Undiscounted Markov Renewal Programming*)

by

P.J. Schweitzer & A. Federgruen

ABSTRACT

This paper investigates the solutions to the functional equations that arise inter alia in Undiscounted Markov Renewal Programming. We show that the solution set is a connected, though possibly non-convex set whose members are unique up to n^* constants, characterize n^* and show that some of these n^* degrees of freedom are locally rather than globally independent.

Our results generalize those obtained in ROMANOVSKY [20] where another approach is followed for a special class of discrete time Markov Decision Processes. Basically our methods involve the set of randomized policies. We first study the sets of pure and randomized maximal-gain policies, as well as the set of states that are recurrent under some maximal-gain policy.

KEY WORDS & PHRASES: *Markov Renewal Programs, average return optimality, functional equations, fixed points*

---

*) This report will be submitted for publication elsewhere

*) Revised version of report BW 60/76

1. INTRODUCTION

This paper investigates the solutions (g,v) to the 2N functional equations:

$$(1.1)\qquad g_i = \max_{k\in K(i)} \sum_{j=1}^{N} P_{ij}^k g_j, \qquad i = 1,\dots,N$$

$$(1.2)\qquad v_i = \max_{k\in L(i)} \left[q_i^k - \sum_{j=1}^{N} H_{ij}^k g_j + \sum_{j=1}^{N} P_{ij}^k v_j\right], \qquad i = 1,\dots,N$$

where

$$(1.3)\qquad L(i) = \left\{k \in K(i) \mid g_i = \sum_{j=1}^{N} P_{ij}^k g_j\right\}.$$

The K(i) are given finite sets and the $q_i^k, P_{ij}^k, H_{ij}^k$ are given arrays with $P_{ij}^k, H_{ij}^k \geq 0$ for all i,j,k; $\sum_{j=1}^N P_{ij}^k = 1$ and $\sum_{j=1}^N H_{ij}^k = T_i^k > 0$, for all i,k. Also we assume property A to be stated below.

For the special cases $H_{ij}^k = P_{ij}^k . \tau_{ij}^k$ with $\tau_{ij}^k \geq 0$ and $H_{ij}^k = \delta_{ij}$, the functional equations arise in Markov Decision Theory with $\Omega = \{1,\dots,N\}$ as state space, q_i^k as the one-step expected reward, P_{ij}^k the transition probability to state j and T_i^k the expected holding time, when alternative k is chosen in state i (cf. BELLMAN [2,3], BLACKWELL [4], HOWARD [11,12], DE CANI [6], JEWELL [13], DENARDO & FOX [8], DENARDO [7], DERMAN [9], SCHWEITZER [21,22,23]). The solution to (1.1) and (1.2) is not unique, although g is uniquely determined. The purpose of this paper is to characterize

$$V = \{v \in E^N \mid v \text{ satisfies } (1.2)\}.$$

We show that V is a connected, though possibly non-convex set whose members are unique up to n^* constants, characterize n^*, and show that some of these n^* degrees of freedom are locally rather than globally independent.

Our results generalize those obtained in ROMANOVSKY [20] where another approach is followed for a special class of discrete time Markov Decision Processes (MDP's).

Basically our methods involve the set of randomized policies. We first study the sets S_{PMG} and S_{RMG} of pure and randomized maximal-gain policies, and characterize the set R^* of states that are recurrent under some maximal gain policy. In section 2 we give the notations and some preliminaries. In section 3 we characterize the sets S_{RMG} and R^*. The properties of V are studied in section 4, while in section 5 the n^* degrees of freedom are characterized. Finally, in section 6 a triangular decomposition of the set V is given.

II. NOTATIONS AND PRELIMINARIES

A *(stationary) randomized policy* f is a tableau $[f_{ik}]$satisfying $f_{ik} \geq 0$ and $\sum_{k \in K(i)} f_{ik} = 1$ for all $i \in \Omega$. In the Markov decision model, f_{ik} denotes the probability that the k^{th} alternative is chosen when entering state i.

We let S_R denote the set of all randomized policies and S_P the subset of all *pure* (non-randomized) policies, i.e. for $f \in S_P$, each $f_{ik} = 0$ or 1 For $f \in S_P$, we use the notation $f^{\#} = (\beta_1,\ldots,\beta_N)$ where $\beta_i \in K(i)$ denotes the single alternative used in state i.

Associated with each $f \in S_R$ are N-component "reward" vector q(f) and "holding time" vector T(f), and two matrices P(f) and H(f):

$$q(f)_i = \sum_{k \in K(i)} f_{ik} q_i^k; \qquad T(f)_i = \sum_{k \in K(i)} f_{ik} T_i^k$$

$$P(f)_{ij} = \sum_{k \in K(i)} f_{ik} P_{ij}^k; \qquad H(f)_{ij} = \sum_{k \in K(i)} f_{ik} H_{ij}^k.$$

Note that P(f) is a stochastic matrix. For any $f \in S_R$, define the stochastic matrix $\Pi(f)$ as the Cesaro limit of the sequence $\{P^n(f)\}_{n=1}^{\infty}$ and define the *fundamental* matrix Z(f) as $[I - P(f) + \Pi(f)]^{-1}$. These matrices always exist and have the following properties (cf. [4],[14]):

$$(2.1) \qquad \Pi(f) = P(f)\Pi(f) = \Pi(f)P(f) = \Pi(f)^2 = \Pi(f)Z(f) = Z(f)\Pi(f)$$

$$(2.2) \qquad [I - P(f)]Z(f) = Z(f)[I - P(f)] = I - \Pi(f)$$

$$(2.3) \qquad Z(f) = I + \lim_{a \uparrow 1} \sum_{n=1}^{\infty} a^n [P(f)^n - \Pi(f)].$$

Denote by $n(f)$ the number of subchains (closed, irreducible sets of states) for $P(f)$. Then:

$$(2.4) \qquad \Pi(f)_{ij} = \sum_{m=1}^{n(f)} \phi_i^m(f)\pi_j^m(f), \qquad 1 \le i, j \le N$$

where the row vector $\pi^m(f)$ is the unique equilibrium distribution of $P(f)$ on the m^{th} subchain $C^m(f)$, and $\phi_i^m(f)$ is the probability of absorption in $C^m(f)$, starting from state i (cf. [7] and [23]). Observe $\sum_i \pi_i^m(f) = 1$ and $\pi^m(f)P(f) = \pi^m(f)$.

Let $R(f) = \{j \mid \Pi(f)_{ij} > 0\}$, i.e. $R(f)$ is the set of recurrent states for $P(f)$. Note that the column vector $\phi^m(f) = P(f)\phi^m(f)$ for all m and that the $\{\phi^m(f) \mid m = 1,\dots,n(f)\}$ are linearly independent. Since any solution to $P(f)x = x$ satisfies $\Pi(f)x = x$ and the rank of $[I - \Pi(f)]$ is $N - n(f)$, it easily follows that the solution set of $P(f)x = x$ is given by:

$$(2.5) \qquad x = \sum_{m=1}^{n(f)} a_m \phi^m(f)$$

with $a_1,\dots,a_{n(f)}$ arbitrary scalars.

LEMMA 2.1. *Fix* $f \in S_R$, *and let the vector* b *satify* $\Pi(f)b = 0$. *Then* $[I - P(f)]x \ge b$, *implies* $x \ge Z(f)b + \Pi(f)x$, *where in both inequalities the equality sign holds for each component* $i \in R(f)$.

PROOF. Multiplying $[I - P(f)]x - b \ge 0$ by $\Pi(f) \ge 0$, yields $0 = \Pi(f)([I - P(f)]x - b)$, implying that the former inequality is a strict equality for components $i \in R(f)$. Using this and the fact that as a result of (2.3), for $j \notin R(f)$, $Z(f)_{ij} \ge 0$ for all i, with $Z(f)_{ij} = 0$ when $i \in R(f)$, we get the desired result by multiplying $[I - P(f)]x \ge b$ by $Z(f)$ and invoking (2.2). □

LEMMA 2.2. *Let* $f \in S_R$, *and let* $C^m(f)$ *be any subchain of* $P(f)$. *Take any* $i \in C^m(f)$ *and any* $k \in K(i)$ *with* $f_{ik} > 0$. *Then there exists a pure policy* h *with* (a) $h_{ik} = 1$, (b) *for every* (j,r), $h_{jr} = 1$ *only if* $f_{jr} > 0$, *such that* (c) i *belongs to a subchain* C *of* $P(h)$ *which is contained within* $C^m(f)$ *and* (d) $R(h) \subseteq R(f)$.

PROOF. Let h meet conditions (a) and (b). Note that, in view of (b), $C^m(f)$ is closed for $P(h)$, since it is closed for $P(f)$. If $C^m(f) = \{i\}$, condition (c) is satisfied. Otherwise, let Δ initially be equal to $\{i\}$. Define $\bar{\Delta} = C^m(f)\backslash\Delta$. Next the following step is performed:

Choose a state $j \in \bar{\Delta}$ and an alternative r such that $f_{jr} > 0$ and $P^r_{jt} > 0$ for some $t \in \Delta$, transfer j from $\bar{\Delta}$ to Δ, and define $h_{jr} = 1$. Clearly, such a j and r can be found, since all states in $C^m(f)$ communicate under $P(f)$. Repeat this step for the new Δ and $\bar{\Delta}$, until $\bar{\Delta}$ is empty. This construction shows that under policy h, state i can be reached from any state in $C^m(f)\backslash\{i\}$. Together this and the fact that $C^m(f)$ is closed under $P(h)$ imply *condition* (c). *Condition* (d) trivially holds if $\Omega = R(f)$. Otherwise, let Γ initially be equal to $R(f)$ and define $\bar{\Gamma} = \Omega - \Gamma$. Choose a state $t_0 \in \bar{\Gamma}$ and a path $\{t_0,t_1,\ldots,t_n\}$ such that $P(f)_{t_\ell t_{\ell+1}} > 0$ for $\ell = 0,\ldots$ $\ldots,n-1$ and $t_n \in \Gamma$. Such a path clearly exists, since t_0 is transient under $P(f)$ and $\Gamma \supseteq R(f)$. Transfer $\{t_0,\ldots,t_{n-1}\}$ from $\bar{\Gamma}$ to Γ and define for $\ell = 0,\ldots,n-1$ $h_{t_\ell r} = 1$ for any r with $f_{t_\ell r} > 0$ and $P^r_{t_\ell t_{\ell+1}} > 0$. Repeat this step until $\bar{\Gamma}$ is empty. Finally, for $j \in R(f) - C^m(f)$, define $h_{jr} = 1$ for some r, with $f_{jr} > 0$ and observe that *condition* (b) holds for all $j \in \Omega$. This completes the proof. □

In the remainder of the paper, we assume that property A holds.

A: If f is any pure policy and $C^m(f)$ is any subchain of $P(f)$, then $i \in C^m(f)$ implies $H(f)_{ij} = 0$ for $j \notin C^m(f)$.

This property is satisfied for both the Markov Renewal Programs (MRP's) with $H^k_{ij} = P^k_{ij}\tau^k_{ij}$ and the discrete time model with $H^k_{ij} = \delta_{ij}$. Using the previous lemma, one easily verifies that if property A holds for all pure policies, it holds for all randomized policies.

LEMMA 2.3. *(Gain and Relative Value Vectors).*
Fix $f \in S_R$. *The general solution to the equations*

(2.6) (a) $g = P(f)g$, (b) $v = q(f) - H(f)g + P(f)v$

is given by

(2.7) $$g_i = g(f)_i = \sum_{m=1}^{n(f)} \phi_i^m(f)g^m(f),$$

with

$$g^m(f) = \langle\pi^m(f),q(f)\rangle/\langle\pi^m(f),T(f)\rangle$$

and

(2.8) $$v_i = Z(f)[q(f) - H(f)g]_i + \sum_{m=1}^{n(f)} a_m\phi_i^m(f),$$

with $a_1,\ldots a_{n(f)}$ *arbitrary scalars.*

PROOF. Note that multiplication of (2.6)(b) by $\Pi(f)$ leads to :

(2.9) $$\Pi(f)[q(f) - H(f)g] = 0.$$

Using property A, it follows from the proof of lemma 1 of [7] that $g(f)$ is the unique solution to (2.6)(a) and (2.9). Hence, any solution (g,v) to (2.6) has $g = g(f)$. Using (2.2) one next verifies by mere insertion that $(g=g(f), v=Z(f)[q(f)-H(f)g(f)])$ satisfy (2.6). Finally (2.8) follows from (2.5), since (2.6)(b) is a linear system of equations with $Z(f)[q(f) - H(f)g(f)]$ as a particular solution. □

The unique solution $g(f)$ to (2.6) will be called the *gain rate vector*, and $g^m(f)$ the gain rate of the subchain $C^m(f)$. A solution v to (2.6) will be called a *relative-value vector* and denoted by $v(f)$.

In the remainder, we will refer to the following example:

EXAMPLE 1. $N = 4$, $K(1) = K(2) = \{1\}$; $K(3) = K(4) = \{1,2\}$; $H_{ij}^k = \delta_{ij}$ for all i,j,k.

i	k	P_{i1}^k	P_{i2}^k	P_{i3}^k	P_{i4}^k	q_i^k
1	1	0	1	0	0	0
2	1	1	0	0	0	0
3	1	1	0	0	0	$q_3^1<0$
3	2	0	0	1	0	0
4	1	.4	.4	.2	0	0
4	2	.8	.2	0	0	0

Using (3.1) and theorem 3.1. part (c) one verifies that

$$V = \{v^* \in E^4 \mid v_1^* = v_2^*;\ v_3^* \geq q_3^1 + v_1^*;\ v_4^* = \max[.8v_1^* + .2v_3^*;\ v_1^*]\}$$

Observe that V is non-convex. Note furthermore, that for $f \in S_{RMG}$, if f makes "unwise" decisions in states in $\Omega - R(f)$, then there do not necessarily exist additive constants such that $v(f) \in V$ (cf. theorem 3 of [22],[25] and our theorem 4.1 part (b)). Take the above example with pure policy $f^{\#} = (1,1,1,1)$ with P(f) unichained, and $v(f) = (0\ 0\ q_3^1\ .2q_3^1) + a(1\ 1\ 1\ 1)$ $\notin V$ for any choice of the additive constant a.

Finally, reference [25] provides examples where the choice of additive constants in v(f) affects the Policy Iteration Algorithm (PIA) (cf. [6],[8],[13]).

III. PROPERTIES OF MAXIMAL GAIN POLICIES

In this section we give some properties of maximal gain policies; some of the notions and properties presented here are related to results in [15],[16],[17],[18].
First, define the *maximal gain rate*

$$(3.1) \qquad g_i^* = \sup_{f \in S_R} g(f)_i, \qquad i = 1,\dots,N.$$

For any $v \in V$, $k \in K(i)$, and $f \in S_R$, define

$$b(v)_i^k = q_i^k - \sum_j H_{ij}^k g_j^* + \sum_j P_{ij}^k v_j - v_i, \qquad i = 1,\dots,N$$

and

$$b(v,f)_i = \sum_{k \in K(i)} f_{ik} b(v)_i^k = [q(f) - H(f)g^* + P(f)v - v]_i \quad i = 1,\dots N.$$

Since g(f) can be interpreted as the average reward of f for a MRP with transition probabilities P_{ij}^k, one-step expected rewards q_i^k, and holding times T_i^k, we know from DERMAN [9] that there exists a pure policy that attains the N suprema in (3.1) simultaneously. Hence $g_i^* = \max_{f \in S_P} g(f)_i$.
Accordingly define:

$$S_{PMG} \quad \{f \in S_P \mid g(f) = g^*\}$$

and

$$S_{RMG} = \{f \in S_R \mid g(f) = g^*\}.$$

Finally, let:

$$w_i^* = \max_{f \in S_{PMG}} Z(f)[q(f) - H(f)g^*]_i.$$

THEOREM 3.1. *(Properties of Maximal-Gain Policies).*

(a) $f \in S_{RMG}$ *if and only if* $g^* = P(f)g^*$ *and* $\Pi(f)[q(f) - H(f)g^*] = 0$.

(b) *The functional equations* (1.1) *and* (1.2) *always have the solution* $g = g^*$, $v = w^*$. *Hence* V *is non-empty. Also, there exists a policy* $f \in S_{PMG}$ *such that* $w^* = Z(f)[q(f) - H(f)g^*]$.

(c) *In any solution* (g,v) *of the functional equations* (1.1) *and* (1.2) $g = g^*$, *hence* g *and each* L(i) *is unique.*

(d) *If* f *is any policy, and if* C *is any subchain of* P(f) *then* $g_i^* =$ *constant,* $i \in C$.

(e) (cf. [15], p. 16, remark 2). *If* $v \in V$, *then* $\max_{k \in L(i)} b(v)_i^k = 0$, *for every* i.

Let $f \in S_R$.

(1) *Suppose that* $k \in L(i)$ *for each* (i,k) *with* $f_{ik} > 0$ *and that for some* $v \in V$, $b(v)_i^k = 0$ *for each* (i,k) *with* $i \in R(f)$ *and* $f_{ik} > 0$. *Then* $f \in S_{RMG}$.

(2) *Conversely, if* $f \in S_{RMG}$, *then for each* $i = 1,\ldots,N$, $f_{ik} > 0$ *implies* $k \in L(i)$, *and for* $i \in R(f)$, $f_{ik} > 0$ *implies* $b(v)_i^k = 0$ *for all* $v \in V$.

PROOF.

(a) As noted in the proof of lemma 2.3, g(f) is the unique solution to the equations $g = P(f)g$ and (2.9).

(b) Invoking the above mentioned interpretation of g^*, we know from theorem 1 in DENARDO & FOX [8] that $g_i^* = \max_k \sum_j P_{ij}^k g_j^*$. Consider the discrete time decision model with $\bar{K}(i) = L(i) = \{k \mid g_i^* = \sum_j P_{ij}^k g_j^*\}$, $\bar{P}_{ij}^k = P_{ij}^k$ and $\bar{q}_i^k = q_i^k - \sum_j H_{ij}^k g_j^*$.

Note that in this model, each policy has $\bar{g}(f) \le 0$. Moreover, it follows from part (a) that $\bar{g}(f) = 0$ if and only if $f \in S_{RMG}$. Hence the discrete time model has $\bar{g}^* = 0$ and, with $\bar{S}_{PMG} =$ $= \{f \in X_{i=1}^N \bar{K}(i) \mid \bar{g}(f) = \bar{g}^* = 0\}$, we have:

$$\max_{f \in S_{PMG}} Z(f)[q(f) - H(f)g^*]_i = \max_{f \in \bar{S}_{PMG}} Z(f)[\bar{q}(f) - \bar{g}^*]_i.$$

for $i = 1,\dots,N$.

Use theorem 4 of [4] in order to prove the existence of a policy $f \in S_{PMG}$ for which $w^* = Z(f)[q(f) - H(f)g^*]$ as well as the fact that w^* satisfies (1.2).

(c) Fix a solution (g,v) to (1.1) and (1.2). Using property A, a minor modification of the proof of lemma 4 of [8], shows that $g \geq g(f)$ for all $f \in S_P$ with equality for any f^o, such that $f^o_{ik} = 1$ for some k maximizing (1.1) and (1.2). Hence $g = g^*$.

(d) Since g^* satisfies (1.1), we have $P(f)g^* \leq g^*$ for all $f \in S_R$. The assertion then follows from lemma 2-a in [8].

(e) The first result follows from the very definition of $b(v)^k_i$

(1) From the definition of $b(v)^k_i$, we have $v_i - \sum_j P(f)_{ij}v_j =$ $q(f)_i - \sum_j H(f)_{ij}g^*_j$ for $i \in R(f)$. Multiplying this equation with $\Pi(f)_{ki}$ and summing over i, we obtain $\Pi(f)[q(f) - H(f)g^*] = 0$. Use this and $g^* = P(f)g^*$ in order to apply part (a).

(2) If $f \in S_{RMG}$, $g^* = P(f)g^*$ follows from part (a). Hence $f_{ik} > 0$ implies $k \in L(i)$ and $b(v)^k_i \leq 0$. So $b(v,f) \leq 0$, for any $v \in V$. Since we know from part (a) that $\Pi(f)b(v,f) = 0$ for $f \in S_{RMG}$, it follows that for $j \in R(f)$, $b(v,f)_j = 0$, i.e. $f_{jk} > 0$ implies $b(v)^k_j \doteq 0$. □

Define next

(3.2) $$R^* = \{i \mid i \in R(f) \text{ for some policy } f \in S_{RMG}\}.$$

The following theorem gives a characterization of this set, which plays a basic part in the remainder of this paper.

<u>THEOREM 3.2</u>. *(Characterization of R^*).*

(a) $R^* = \{i \mid i \in R(f)$ *for some* $f \in S_{PMG}\}$.

(b) *The set* $\{f \in S_{RMG} \mid R(f) = R^*\}$ *is not empty.*

(c) *Define* $n^* = \min\{n(f) \mid f \in S_{RMG}$ *with* $R(f) = R^*\}$ *and* $S^*_{RMG} =$ $= \{f \in S_{RMG} \mid R(f) = R^*$ *and* $n(f) = n^*\}$. *Fix* $f^* \in S^*_{RMG}$. *Any subchain of any* $f \in S_{RMG}$ *is contained within a subchain of* $P(f^*)$.

(d) *All* $f^* \in S^*_{RMG}$ *have the same collection of subchains* $\{R^{*\alpha}, \alpha = 1,\dots,n^*\}$.

(e) *For any* α, $1 \le \alpha \le n^*$, $g^*_i = g^{*\alpha}$ *(say) for all* $i \in R^{*\alpha}$.

(f) *Let* $R^{(1)},\dots,R^{(m)}$ *be disjoint sets of states such that*

(1) *if* C *is a subchain of some* $f \in S_{RMG}$, *then* $C \subseteq R^{(k)}$ *for some* k, $1 \le k \le m$;

(2) *there exists a* $f^* \in S_{RMG}$ *with* m *subchains* $\{R^{(k)}\}^m_{k=1}$.

Then $m = n^*$ *and after renumbering* $R^{(\alpha)} = R^{*\alpha}$ *for* $\alpha = 1,\dots,n^*$.

PROOF.

(a) Fix a state $i \in R^*$ and a $f \in S_{RMG}$ such that $i \in R(f)$. Consider a policy h satisfying the conditions (a), (b), (c) and (d) of lemma 2.2. Then $i \in R(h)$, whereas $h \in S_{PMG}$ is verified with the help of theorem 3.1 part (e). Thus the right-hand side of (a) includes R^* and the reversed inclusion is immediate.

(b) Fix an enumeration $f^1,\dots,f^M$ of S_{PMG}. For any $i \in R^*$, let $A_i = \{r \mid i \in R(f^r)\}$. Consider the following equivalence relation on $\mathcal{C} \div \{C^m(f^r) \mid 1 \le r \le M;\ 1 \le m \le n(f^r)\}$:

Let $C \sim C'$, if there exists $\{C^{(1)}=C, C^{(2)},\dots,C^{(n)}=C'\}$ with $C^{(i)} \in \mathcal{C}$ and $C^{(i)} \cap C^{(i+1)} \ne \emptyset$ for $i = 1,\dots,n-1$.

Let f^* satisfy: (1) $\{k \mid f^*_{ik} > 0\} = \cup_{r\in A_i} \{k \mid f^r_{ik} > 0\}$ for $i \in R^*$; (2) $\{k \mid f^*_{ik} > 0\} = L(i)$ for $i \in \Omega - R^*$. Using theorem 3.1 part (e) one verifies that $f^* \in S_{RMG}$.

Clearly, the equivalence classes are the subchains of $P(f^*)$ since they are closed under $P(f^*)$ and since the states belonging to the same equivalence class communicate with each other. Hence, $R^* = R(f^*)$.

(c) Assume $P(f)$, for $f \in S_{RMG}$, has a subchain $C^m(f)$ that intersects say the subchains R^{*1} and R^{*2} of $P(f^*)$. Then a policy f^{**} with $\{k \mid f^{**}_{ik} > 0\} = \{k \mid f^*_{ik} > 0\} \cup \{k \mid f_{ik} > 0\}$ for all $i \in C^m(f)$, and $\{k \mid f^{**}_{ik} > 0\} = \{k \mid f^*_{ik} > 0\}$ otherwise, is maximal gain, has $R(f^{**}) = R^*$, and its number of subchains is at most $n^* - 1$, since the states of R^{*1} and R^{*2} communicate with each other under $P(f^{**})$. This contradicts the minimality of n^*.

(d) For all f^*, $f^{**} \in S^*_{RMG}$, part (c) implies each $C^\alpha(f^*) \subseteq$ some $C^\beta(f^{**})$, and each $C^\beta(f^{**}) \subseteq C^\alpha(f^*)$.

(e) Combine part (d) with part (d) of theorem 3.1.

(f) Apply property (1) to conclude $R^{*\alpha} \subseteq R^{(k(\alpha))}$. Apply part (c) and property (2) to conclude $R^{(k(\alpha))} \subseteq R^{*\alpha}$. □

COROLLARY 3-3:

Let policy f^* *be constructed as in the proof of part* (b) *of the previous theorem. Then,* $f^* \in S^*_{RMG}$.

PROOF. Let $R^{(1)},\dots,R^{(m)}$ be the subchains of $P(f^*)$. Fix a subchain C of P(f), for some $f \in S_{RMG}$. Recall from part (b) of th. 3.2 that $C \subset R^* = \cup_{k=1}^M R^{(k)}$, and let C intersect $R^{(1)}$ (say). We prove the corollary by verifying that $C \subseteq R^{(1)}$ (cf. part (f) of th.3.2) which in turn follows by showing that for any $i \in C \cap R^{(1)}$:
$\{j \mid P(f)_{ij} > 0\} = \cup_{\{k \mid f_{ik}>0\}} \{j \mid P^k_{ij} > 0\} \subset C \cap R^{(1)}$. Fix $i \in C \cap R^{(1)}$ and k, such that $f_{ik} > 0$. Apply lemma 2.2 in combination with th. 3.1. part (e) to verify the existence of a pure policy $h \in S_{PMG}$, with $i \in R(h)$ and $h_{ik} = 1$ and conclude from the definition of policy f^* that
$\{j \mid P^k_{ij} > 0\} \subset R^{(1)} \cap C$. □

Note that randomization, by coalescing subchains, is essential for the recurrency properties: in general, there may fail to exist a pure maximal gain policy f with $R(f) = R^*$, or which achieves the minimal number n^* of subchains.

A finite procedure for calculating R^*, n^*, the $R^{*\alpha}$ and a $f^* \in S^*_{RMG}$ is therefore as follows: use the PIA to find g^* and a $v \in V$. Compute $S_P(v) = X_{i=1}^N \{k \in L(i) \mid b(v)_i^k = 0\} = \{f \in S_P \mid f$ achieves the 2N minima in (1.1) and (1.2)$\} \subseteq S_{PMG}$. Parts (a) of th. 3.1 and th. 3.2 together establish $R^* = \{i \mid i \in R(f)$, for some $f \in S_P(v)\}$ (cf. also [17], algorithm on p. 353-4). Determine $\{R^{*\alpha}\}$ as the equivalence classes of the set of subchains of policies belonging to $S_P(v)$ (cf. proof of theorem 3.1. part (b) and corollary 3.3). Finally, define f^* by $\{k \mid f^*_{ik} > 0\} = L(i)$ for $i \in \Omega \setminus R^*$, and $\{k \mid f^*_{ik} > 0\} = \{k \in L(i) \mid b(v)^k_i = 0, \sum_{j \in R^{*\alpha}} P^k_{ij} = 1\}$ for $i \in R^{*\alpha}$ ($\alpha = 1,\ldots,n^*$).

IV. PROPERTIES OF V

Some basic properties of V are given by:

THEOREM 4.1. *(Basic Properties of* V*).*

(a) V *is closed and unbounded, as* $v \in V$ *implies* $v + a_1\underline{1} + a_2 g^* \in V$, *for any scalars* a_1, a_2 *(where* $\underline{1}$ *is the* N*-vector with all coordinates unity).*

(b) *(Maximality of relative values.) For any* $v^* \in V$ *and* $f \in S_{RMG}$, *it is possible to choose the* $n(f)$ *additive constants in* $v(f)$ *such that* $v^* \geq v(f)$ *with equality for components in* $R(f)$.

(c) *(Cf.* [3],[15],[16],[21]) $v \in V$, *if and only if*

$$v_i = \max_{f \in S_{PMG}} \{Z(f)[q(f) - H(f)g^*]_i + \Pi(f)v_i\} \qquad i = 1,\ldots,N. \tag{4.1}$$

In addition, if $v \in V$, *then a policy* $f \in S_{PMG}$ *achieves all* N *maxima in* (4.1) *if and only if it achieves the* 2N *maxima in* (1.1) *and* (1.2).

PROOF.

(a) Immediate to verify.

(b) Choose in (2.8) $a_m = \langle \pi^m(f), v^* \rangle$. From part (e) of theorem 3.1, it follows that $\{k \mid f_{ik} > 0\} \subseteq L(i)$ for each i, hence $v^* \geq q(f) - H(f)g^* + P(f)v^*$, which implies, using th. 3.1 part (a), lemma 2.1, (2.4) and (2.8):

$$v^* \geq Z(f)[q(f) - H(f)g^*] + \Pi(f)v^* =$$

$$= Z(f)[q(f) - H(f)g^*] + \sum_{m=1}^{n(f)} a_m \phi^m(f) = v(f)$$

with equality for components in R(f).

(c) First assume $v \in V$. In part (b) we proved that for any $f \in S_{PMG}$, $v \geq Z(f)[q(f) - H(f)g^*] + \Pi(f)v$, with strict equality for $f \in S_P(v)$. Hence, $v \in V$ implies (4.1) and any policy achieving the 2N maxima in (1.1) and (1.2) achieves all N maxima in (4.1).

Conversely, if v satisfies (4.1), we define

(4.2) $$\tilde{v}_i = \max_{k\in L(i)} [q_i^k - \sum_j H_{ij}^k g_j^* + \sum_j P_{ij}^k v_j], \qquad i = 1,\dots,N$$

and show both $\tilde{v} \ge v$ and $\tilde{v} \le v$, hence $\tilde{v} = v \in V$.

For any $f \in S_{PMG}$, $f_{ik} = 1$ implies $k \in L(i)$ by theorem 3.1 part (e); hence using (4.1), (2.2) and th. 3.1 part (a):

$$\tilde{v} \ge q(f) - H(f)g^* + P(f)v \ge [I + P(f)Z(f)][q(f) - H(f)g^*]+\Pi(f)v =$$

$$= Z(f)[q(f) - H(f)g^*] + \Pi(f)v, \qquad f \in S_{PMG}.$$

This implies $\tilde{v} \ge v$. Let h denote a pure policy in $X_{i=1}^N L(i)$, achieving all maxima in (4.2). Then:

(4.3) $$v_i \le \tilde{v}_i = [q(h) - H(h)g^* + P(h)v]_i; \qquad i = 1,\dots,N$$

Multiply (4.3) with $\Pi(h) \ge 0$ in order to get $0 \le \Pi(h)[q(h) - H(h)g^*] \le 0$, the latter inequality following from (2.9) and $g(h) \le g^*$. Hence $h \in S_{PMG}$, by part (a) of th. 3.1.

Using lemma 2.1, (4.3) implies $v \le Z(h)[q(h) - H(h)g^*] + \Pi(h)v$. Insert this on the right-hand side of (4.2) and use $\Pi(h)[q(h) - H(h)g^*] = 0$, to obtain:

$$\tilde{v} \le [I + P(h)Z(h)][q(h) - H(h)g^*] + \Pi(h)v =$$

$$= Z(h)[q(h) - H(h)g^*] + \Pi(h)v \le$$

$$\le \max_{f\in S_{PMG}} \{Z(f)[q(f) - H(f)g^*] + \Pi(f)v\} = v.$$

Finally, if $f \in S_{PMG}$ achieves the N maxima in (4.1), multiply the resulting equality in (4.1) with $Z(f)^{-1}$ to show that it achieves the N maxima in (1.2), as well as the N maxima in (1.1), since $f_{ik} = 1$ implies $k \in L(i)$. This completes the proof. □

Since for $f \in S_{RMG}$, $\Pi(f)_{ij} = 0$ if $j \notin R^*$, we have by part (c) of theorem 4.1 that $v \in V$ if and only if

$$(4.4)\qquad v_i = \max_{f\in S_{PMG}} \{Z(f)[q(f) - H(f)g^*]_i + \sum_{j\in R^*} \Pi(f)_{ij}v_j\}, \qquad i \in R^*$$

$$(4.5)\qquad v_i = \max_{f\in S_{PMG}} \{Z(f)[q(f) - H(f)g^*]_i + \sum_{j\in R^*} \Pi(f)_{ij}v_j\}, \qquad i \in \Omega\backslash R^*.$$

Observe that (4.4) involves only $(v_i|i\in R^*)$ and can be studied in isolation. The $(v_i|i\in\Omega\backslash R^*)$ are uniquely determined via (4.5), for any $(v_i|i\in R^*)$. Define now

$$(4.6)\qquad V^R = \{(v_i|i\in R^*);\ v_i \text{ satisfy (4.4) for all } i \in R^*\}.$$

THEOREM 4.2.

(a)

$$(4.7)\qquad V^R = \{(v_i|i\in R^*);\ v_i \geq Z(f)[q(f) - H(f)g^*]_i + \sum_{j\in R^*} \Pi(f)_{ij}v_j, \textit{ for all } i \in R^*,\ f \in S_{PMG}\}.$$

Hence, V^R *is a closed, convex, unbounded, polyhedral set.*

(b) V *is connected.*

PROOF.

(a) Clearly, V^R is contained within the polyhedron that is defined in the right side of (4.7). Conversely fix $i \in R^*$ and $h \in S_{PMG}$ with $i \in R(h)$ (cf. th. 3.2 part (a)). Then , by multiplying the inequalities in (4.7) with $\Pi(h) \geq 0$, we obtain $v_i = Z(h)[q(h) - H(h)g^*]_i + \sum_{j\in R^*} \Pi(h)_{ij}v_j$; hence (4.4) holds. The unboundedness of V is proved as in th. 4.1.

(b) The assertion follows by showing that for any $v,\tilde{v} \in V$, the curve $\{\underline{v}(\lambda) \mid \lambda \in [0.1]\}$ with parameter representation: $v(\lambda)_i = \lambda v_i + (1-\lambda)\tilde{v}_i$, $i \in R^*$ and $v(\lambda)_i = \max_{f\in S_{PMG}} \{Z(f)[q(f) - H(f)g^*]_i + \sum_{j\in R^*} \Pi(f)_{ij}v(\lambda)_j\}$, for $i \notin R^*$, connects v with $\tilde{v}$, lies within V as a consequence of (4.5) and part (a), and is continuous, since all its components are continuous functions of λ. □

We already saw that V may not be convex. The following theorem gives a necessary and sufficient condition for the convexity of V.
This property is especially important when considering MRPs, where for

several quantities of interest (e.g. the optimal bias vector) variational characterizations may be obtained of the nature: $\max_{v\in V}[c + Bv]$ (where c and B are expressions in q_i^k, P_{ij}^k and H_{ij}^k) and the latter is a linear program if and only if V is convex.

THEOREM 4.3. V *is convex if and only if for each* $i \in \Omega - R^*$ *there exists an alternative* $k(i) \in L(i)$, *such that for all* $v \in V$:

$$(4.8) \qquad v_i = q_i^{k(i)} - \sum_j H_{ij}^{k(i)} g_j^* + \sum_j P_{ij}^{k(i)} v_j.$$

Moreover, V *is convex if and only if it is a polyhedron.*

PROOF. We first observe that for any $i \in R^*$, there is a $h \in S_{PMG}$, with $i \in R(h)$, hence by part (e) of theorem 3.1 there exists an alternative $k(i) \in L(i)$ with $b(v)_i^{k(i)} = 0$, for any $v \in V$. Thus (4.8) always holds for $i \in R^*$. Suppose it holds for $i \in \Omega - R^*$ as well. Then the functional equations (1.2) are equivalent to the linear (in)equalities $b(v)_i^{k(i)} = 0$ for $i = 1,\dots,N$ and $b(v)_i^k \le 0$ for $k \in L(i)\backslash\{k(i)\}$ and $i = 1,\dots,N$. Hence V is a convex polyhedron.

Conversely, suppose V is convex. Assume to the contrary that there exists a state $i \in \Omega - R^*$ and a finite set of $v^{(m)}$'s in V, such that no $k \in L(i)$ achieves the maximum in (1.2) for all $v^{(m)}$. However, since V is convex, it is immediate to verify that a $k \in L(i)$ achieving the maximum in (1.2) for a positive convex combination $\bar{v}$ of the $v^{(m)}$'s, achieves the maximum in (1.2) for each $v^{(m)}$. □

REMARK 1. Condition (4.8), hence convexity of V, holds trivially if (1) $R^* = \Omega$, or (2) L(i) is a singleton for each $i \in \Omega\backslash R^*$, or (3) there is only one maximal gain policy or (4) $n^* = 1$, since $v \in V$ is unique up to a multiple of $\underline{1}$ (cf. remark 2.)

For discrete time Markovian decision processes, where $H_{ij}^k = \delta_{ij}$, the value iteration equations take the form:

$$(4.9) \qquad v(n+1)_i = \max_{f\in K(i)} \{q_i^k + \sum_j P_{ij}^k v(n)_j\},$$

with v(0) a given vector.

It is well known that $\{v(n) - ng^*\}_{n=1}^{\infty}$ may fail to converge. In a forthcoming paper [24] it will be shown that there exists an integer J such that

$$u_i^{(r)} = \lim_{n\to\infty} \{v(nJ+r)_i - (nJ+r)g_i^*\}$$

exists for all i, with $u_i^{(r+J)} = u_i^{(r)}$ (previous proofs in [5] and [15] are both incorrect).

Accordingly, define $\bar{v}$ as the Cesaro-limit of the sequence $\{v(n) - ng^*\}_{n=1}^{\infty}$. Example 1 with v(0) = [1 0 1 .6] shows that in general $\bar{v} \notin V$ ($v(2n)_1=1$; $v(2n+1)_1=0$; $v(2n)_2=0$; $v(2n+1)_2=1$; $v(n)_3=1$; $v(2n)_4=.6$; $v(2n+1)_1=.8$; $\bar{v}=[.5\ .5\ 1\ .7] \notin V$).

The relation between $\bar{v}$ and V is as follows:

THEOREM 4.4.
(a) $\{\bar{v}_i \mid i \in R^*\} \in V^R$.
(b) *There exists a vector* $v \in V$, *such that* $v \leq \bar{v}$ *with equality for components in* R^*.

PROOF. Note that for all $i \in \Omega$: $u_i^{(r+1)} = \max_{k\in K(i)} \{q_i^k - g_i^* + \sum_j P_{ij}^k u_j^{(r)}\}$, since for all n sufficiently large the maximizing alternatives in (4.9) belong to L(i) as observed in [5] and [15].

Since $\bar{v} = \frac{1}{J}\sum_{r=0}^{J-1} u^{(r)}$, we obtain by averaging over r = 0,...,J - 1:

$$\bar{v}_i \geq q_i^k - g_i^* + \sum_j P_{ij}^k \bar{v}_j, \qquad i = 1,\ldots,N \text{ and } k \in K(i).$$

Take any $f \in S_{PMG}$ to obtain: $\bar{v} \geq q(f) - g^* + P(f)\bar{v}$, and hence, using lemma 2.1: $\bar{v} \geq Z(f)[q(f) - g^*] + \Pi(f)\bar{v}$, with equality for $i \in R(f)$. This implies: $\bar{v} \geq \max_{f\in S_{PMG}} \{Z(f)[q(f) - g^*] + \Pi(f)\bar{v}\}$ with equality for components in R^*. Using (4.4) and (4.5) we obtain that the vector v defined by (1) $v_i = \bar{v}_i$, $i \in R^*$ and (2) $v_i = \max_{f\in S_{PMG}} \{Z(f)[q(f) - g^*]_i + \sum_{j\in R^*} \Pi(f)_{ij}\bar{v}_j\}$ for $i \in \Omega - R^*$, belongs to V with $v \leq \bar{v}$ and equality for components in R^*. □

V. THE n^* DEGREES OF FREEDOM IN V

In this section we show that the convex polyhedral set V^R has dimension n^* and that its elements, and hence V, are fully determined by n^* parameters $(y_1,\dots,y_{n^*})$.

ROMANOVSKY [20] obtained the same result for the functional equations that arise in discrete time Markov models with $\underline{g}^* = \langle g^* \rangle \underline{1}$. In addition, as our methods involve the chain structure, a fuller characterization of the parameter space is possible.

The key observation is that any two vectors $v^o, \tilde{v} \in V$ have the property: $\tilde{v}_i - v_i^o = \text{constant} = y_\alpha$ for $i \in R^{*\alpha}$, $\alpha = 1,\dots,n^*$. By fixing $v^o \in V$ and picking these n^* constants, one thus determines $(\tilde{v}_i \mid i \in R^*)$ and hence $\tilde{v}$ by (4.5) in terms of v^o. Hence, by fixing v^o, and sweeping out all permitted values of y, we sweep out all vectors $\tilde{v}$ in V. In particular (5.1) below shows that $\tilde{v}$ is a convex piecewise linear function in y.

THEOREM 5.1. *Let* $v \in V$. *The following are equivalent:*

(a) $v + x \in V$

(b) $x_i = \max_{k \in L(i)} [b(v)_i^k + \sum_j P_{ij}^k x_j]$, $\quad i = 1,\dots,N$

(c) $x_i = \max_{f \in S_{PMG}} [Z(f)b(v,f) + \Pi(f)x]_i$ $\quad i = 1,\dots,N$

(d) *there are* n^* *constants* $y = (y_1,\dots,y_{n^*})$ *satisfying*

$$(5.1) \qquad x_i = \begin{cases} y_\alpha & i \in R^{*\alpha},\ \alpha = 1,\dots,n^* \\ \max\limits_{f \in S_{PMG}} \left[Z(f)b(v,f)_i + \sum\limits_{\beta=1}^{n^*} \left(\sum_{j \in R^{*\beta}} \Pi(f)_{ij} \right) y_\beta \right], & i \in \Omega \setminus R^* \end{cases}$$

$$(5.2) \qquad y_\alpha \ge Z(f)b(v,f)_i + \sum_{\beta=1}^{n^*} \left(\sum_{j \in R^{*\beta}} \Pi(f)_{ij} \right) y_\beta, \qquad \alpha = 1,\dots,n^*;\ i \in R^{*\alpha},\ f \in S_{PMG}.$$

PROOF.

(a) $\iff$ (b): (b) is the requirement that $v + x \in V$.

(a) $\iff$ (c): Cf. (4.1) and the definition of b(v,f).

(a) $\Rightarrow$ (d): Take $f^* \in S^*_{RMG}$. As $v, v + x \in V$, we have from part (e) of theorem 3.1: $v_i = [q(f^*) - H(f^*)g^* + P(f^*)v]_i$ and $(v+x)_i = [q(f^*) - H(f^*)g^* + P(f^*)(v+x)]_i$ for all $i \in R^* = R(f^*)$.

Subtraction yields: $x_i = [P(f^*)x]_i = [\Pi(f^*)x]_i = \langle\pi^\alpha(f^*),x\rangle$ for $i \in R^{*\alpha}$, which proves the first part of (5.1). Moreover, this implies the remainder of (d), using (4.4) and (4.5) and the definition of $b(v,f)$.

(d) $\Rightarrow$ (a): Use (4.4), (4.5) and the definition of $b(v,f)$. □

Fix $v \in V$. Define the set of allowed constants

$$Y(v) = \{y \in E^{n^*} \mid y \text{ satisfies } (5.2)\}.$$

Note that,

$$(5.3) \qquad Z(f)b(v,f) \le 0 \qquad\qquad \text{for all } f \in S_{PMG}.$$

(5.3) follows from lemma 2.1, with $x = 0$, using $b(v,f) \le 0$ and $\Pi(f)b(v,f) = 0$ (cf. theorem 3.1 part (d) and (e)).

Clearly, by (5.3), (5.2) is automatically satisfied for (α,i,f) with $\sum_{j\in R^{*\alpha}} \Pi(f)_{ij} = 1$. We accordingly define:

$$\tilde{K}(\alpha) = \{(i,f) \mid i \in R^{*\alpha},\ f \in S_{PMG},\ \sum_{j\in R^{*\alpha}} \Pi(f)_{ij} < 1\},\ \alpha = 1,\dots,n^*,$$

and make the partition $\{1,2,\dots,n^*\} = E \cup F$, where $E = \{\alpha \mid \tilde{K}(\alpha) = \emptyset\}$, $F = \{\alpha \mid \tilde{K}(\alpha) \ne \emptyset\}$,

For $\xi = (i,f) \in \tilde{K}(\alpha)$, define

$$\tilde{q}_\alpha^\xi = [Z(f)b(v,f)]_i, \qquad \text{and} \qquad \tilde{P}_{\alpha\beta}^\xi = \sum_{j\in R^{*\beta}} \Pi(f)_{ij}.$$

Note that $\tilde{q}_\alpha^\xi \le 0$, $\tilde{P}_{\alpha\beta}^\xi \ge 0$, $\sum_{\beta=1}^{n^*} \tilde{P}_{\alpha\beta}^\xi = 1$, $\tilde{P}_{\alpha\alpha}^\xi < 1$ for all $\alpha \in F$, and $\xi \in \tilde{K}(\alpha)$. Then $Y(v)$ consists of all $y \in E^{n^*}$ satisfying

$$(5.4) \qquad y_\alpha \ge \tilde{q}_\alpha^\xi + \sum_{\beta=1}^{n^*} \tilde{P}_{\alpha\beta}^\xi y_\beta, \qquad\qquad \alpha \in F,\ \xi \in \tilde{K}(\alpha).$$

In order to show that $Y(v)$ is a n^*-dimensional polyhedral set, we need the following discrete time Markovian model with state space $\{1,\dots,n^*\}$: For $\alpha \in F$, let $\tilde{K}(\alpha)$ be the set of feasible decisions. For $\xi \in \tilde{K}(\alpha)$, let $\tilde{q}_\alpha^\xi$ and $\tilde{P}_{\alpha\beta}^\xi$ denote the associated reward and transition probabilities (we already noted that $\tilde{P}^{\xi}_{\alpha\beta} \geq 0$, $\sum_{\beta} \tilde{P}^{\xi}_{\alpha\beta} = 1$).

For $\alpha \in E$, add a decision ξ_o to the empty $\tilde{K}(\alpha)$ with $\tilde{q}^{\xi_o}_{\alpha} = -1$ and $\tilde{P}^{\xi_o}_{\alpha\beta} = \delta_{\alpha\beta}$. Let Φ denote the set of pure policies. For $\varphi \in \Phi$, the quantities $\tilde{q}(\varphi)$, $\tilde{P}(\varphi)$, $\tilde{\Pi}(\varphi)$ and $\tilde{Z}(\varphi)$ are defined analogously to q(f), P(f), Π(f) and Z(f) for $f \in S_P$. Also let $\{\tilde{g}^*_{\alpha}\}$ be the maximal gain vector for the new process. Note that $\tilde{q}(\varphi) \leq 0$ for any $\varphi \in \Phi$, so $\tilde{g}^*_{\alpha} \leq 0$ for all α. Also $\tilde{g}^*_{\alpha} = -1$ for $\alpha \in E$, since each state $\alpha \in E$ is a trapping state for $\tilde{P}(\varphi)$, for all $\varphi \in \Phi$. The following lemma characterizes the subchains of $\tilde{P}(\varphi)$ on F:

LEMMA 5.2. (*Properties of subchains of* $\tilde{P}(\varphi)$ *on* F)
Fix $v \in V$. *Assume* $F \neq \emptyset$. *Suppose for some policy* $\varphi \in \Phi$, $\tilde{P}(\varphi)$ *has a subchain* $C \subset F$. *Then*

(a) C *has at least two numbers*

(b) $\tilde{q}(\varphi)_{\alpha}$ *is strictly negative for at least one* $\alpha \in C$.

PROOF.

(a) Part (a) follows from $\tilde{P}^{\xi}_{\alpha\alpha} < 1$ for any $\alpha \in F$ and $\xi \in \tilde{K}(\alpha)$.

(b) Let policy ϕ use action $(i(\alpha),f(\alpha)) \in \tilde{K}(\alpha)$ for each $\alpha \in C$. For $\alpha \in C$, define $S(\alpha) = \{j \mid P(f(\alpha))^n_{i(\alpha)j} > 0$, for some $n = 0,1,2,\ldots\}$. Note that $i(\alpha) \in S(\alpha)$ and that:

(5.5) $\qquad \alpha \in C$, $i \in S(\alpha)$ imply $P(f(\alpha))_{ij} > 0$ only if $j \in S(\alpha)$.

Now assume to the contrary that for each $\alpha \in C$, $0 = \tilde{q}(\phi)_{\alpha} =$
$= Z(f(\alpha))b(v,f(\alpha))_{i(\alpha)}$. Since $f(\alpha) \in S_{PMG}$, $b(v,f(\alpha)) \leq 0$ with equality for components in R(f(α)). Hence, using (2.3), $0 = \tilde{q}(\phi)_{\alpha} =$
$= \sum_{j \notin R(f(\alpha))} Z(f(\alpha))_{i(\alpha)j}\, b(v,f(\alpha))_j =$
$= \sum_{(j \notin R(f(\alpha))} \sum_{n=0}^{\infty} [P(f(\alpha))]^n_{i(\alpha)j} b(v,f(\alpha))_j$ where the interchange of $\sum_n$ and $\lim_{a\uparrow 1}$ is justified by the monotone convergence theorem.
Hence:

(5.6) $\qquad b(v,f(\alpha))_j = 0 \qquad\qquad$ for $j \in S(\alpha)$, $\alpha \in C$.

We now exhibit a policy $f^o \in S_{RMG}$ with the contadictory properties that $R^o = \cup_{\alpha\in C} [R^{*\alpha} \cup S(\alpha)]$ is closed under $P(f^o)$ while every state in R^o is transient for $P(f^o)$.

Take $f^* \in S^*_{RMG}$. Define f^o as follows:
Initially, for $i \in R^*$ set $\{k \mid f^o_{ik} > 0\} = \{k \mid f^*_{ik} > 0\}$. Then for $i \in S(\alpha)$ add $\{k \mid f(\alpha)_{ik} > 0\}$ to $\{k \mid f^o_{ik} > 0\}$. Finally, for $i \in \Omega \setminus R^o$, set $\{k \mid f^o_{ik} > 0\} = \{k \in L(i) \mid b(v)^k_i = 0\}$.

From (5.6) the definition of f^* in combination with theorem 3.1 part (e), and the definition of f^o on $\Omega \setminus R^o$ it follows that $f^o_{ik} > 0$ implies $b(v)^k_i = 0$, for all i, hence $f^o \in S_{RMG}$.

For $i \in R^o$, (5.5) and the fact that $f^* \in S^*_{RMG}$ imply that $P(f^o)_{ij} > 0$ only for $j \in R^o$; hence, R^o is closed under $P(f^o)$.

As $\sum_{j \notin R^{*\alpha}} \Pi(f(\alpha))_{i(\alpha)j} > 0$, there exists a $j \notin R^{*\alpha}$, and an integer $n \geq 1$, with $P(f(\alpha))^n_{i(\alpha)j} > 0$ and so $P(f^o)^n_{i(\alpha)j} > 0$. Hence $i(\alpha) \in R^{*\alpha}$ is transient under $P(f^o)$, since the subchains of a maximal gain policy are all contained within a single $R^{*\beta}$ (cf. theorem 3.2 part (c)).

Now, observe that for each $\alpha \in C$, all states in $R^{*\alpha}$ communicate with $i(\alpha) \in R^{*\alpha}$ for $P(f^o)$, since they communicate with $i(\alpha)$ for $P(f^*)$. However, this implies that each state in $\cup_{\alpha \in C} R^{*\alpha}$ is transient, since a transient state cannot be reached from a recurrent state.

It remains to prove that each $j \in S(\alpha)$, $(\alpha \in C)$, is transient for $P(f^o)$. Fix $j \in S(\alpha)$, $\alpha \in C$. Since $f(\alpha)$ is maximal gain, there is a state $r \in R^{*\beta}$, for some β, such that $P(f(\alpha))^m_{jr} > 0$, for some $m \geq 1$. Hence $P(f^o)^m_{jr} > 0$. Let n be such that $P(f(\alpha))^n_{i(\alpha)j} > 0$. Finally $\beta \in C$ follows from

$$\tilde{P}(\phi)_{\alpha\beta} \geq \Pi(f(\alpha))_{i(\alpha)r} = [P(f(\alpha))^n \Pi(f(\alpha))]_{i(\alpha)r} \geq$$

$$\geq P(f(\alpha))^n_{i(\alpha)j}\, \Pi(f(\alpha))_{jr} > 0$$

and the fact that C is a subchain of $\tilde{P}(\phi)$. This implies that r is transient for $P(f^o)$ and so is j, since a transient state cannot be reached from a recurrent state. □

Together part (b) of lemma 5.2 and the choice of $\tilde{q}^{\xi}_{\alpha}{}^{o} = -1$ for $\alpha \in E$ imply:

(5.7) $\qquad \tilde{g}^*_\alpha < 0 \qquad$ for $\qquad \alpha = 1,\dots,n^*$.

THEOREM 5.3. (*cf. theorem* 3 *of* [20]). *Fix* $v \in V$. *Given any* $\{y_\alpha \mid \alpha \in E\}$ *there exist* $\{y_\alpha \mid \alpha \in F\}$ *such that the following strict inequalities hold:*

$$(5.8) \qquad y_\alpha > \tilde{q}_\alpha^\xi + \sum_{\beta=1}^{n^*} \tilde{P}_{\alpha\beta}^\xi y_\beta, \qquad \text{for all} \quad \alpha \in F,\ \xi \in \tilde{K}(\alpha)$$

PROOF. It suffices to show that there exists a solution y^o to (5.8) for some $\{y_\alpha^o \mid \alpha \in E\}$ since a solution for any $\{y_\alpha \mid \alpha \in E\}$ is then obtained by first adding a large positive constant to every y_α, and then reducing $\{y_\alpha \mid \alpha \in E\}$ to the desired magnitudes, thereby strengthening the inequalities (5.8).

Since $\tilde{q}_\alpha^{\xi_o} = -1$ and $\tilde{P}_{\alpha\alpha}^{\xi_o} = 1$, for $\alpha \in E$, the solution set to (5.8) is not altered by adding the inequalities $y_\alpha > \tilde{q}_\alpha^{\xi_o} + \sum_{\beta=1}^{n^*} \tilde{P}_{\alpha\beta}^{\xi_o} y_\beta$, $\alpha \in E$. Now assume to the contrary, that the solution set of (5.8) is empty. Then for the LP-problem:

$$\min Z \qquad \text{subject to}$$

$$y_\alpha + Z \geq \tilde{q}_\alpha^\xi + \sum_{\beta=1}^{n^*} \tilde{P}_{\alpha\beta}^\xi y_\beta, \qquad \alpha = 1,\ldots,n^*;\ \xi \in \tilde{K}(\alpha),$$

we have $\min Z \geq 0$, which according to theorem 2 of [19], implies $\max_{\alpha=1,\ldots,n^*} \tilde{g}_\alpha^* \geq 0$. This contradicts (5.7). □

Since the solution set to (5.8) is open, for any y satisfying (5.8), there exists a $\delta > 0$, so that $|y - y'| < \delta$ implies $y' \in Y(v)$. Hence the n^* parameters $(y_1,\ldots,y_{n^*})$ may be chosen independently over some (finite) region. V and V^R have exactly $n^* = |E \cup F|$ degrees of freedom of which $|E|$ are globally independent and $|F|$ are only locally independent. Examples can be constructed where E (or F) can be empty; e.g. F is empty if $n^* = 1$. Finally note:

REMARK 2. $n^* = 1 \iff v \in V$ is unique up to a multiple of $\underline{1}$.

VI. PROPERTIES of Y(v): A TRIANGULAR DECOMPOSITION

The parameter set Y(v), for any $v \in V$, is given by (5.2) (or (5.4)) and possesses a canonical representation as in [10].

Here we are able to give a more extensive triangular decomposition of $Y(v)$, based upon a classification of the state space $\tilde{\Omega} = \{1,\dots,n^*\} = E \cup F$ which is related to the one described in BATHER [1].

The decomposition employs the following notions:
State $\alpha \in \tilde{\Omega}$ is said to *have access* to state $\beta \in \tilde{\Omega}$ if there exists a policy $\varphi \in \Phi$ such that $\tilde{P}(\varphi)^m_{\alpha\beta} > 0$ for some integer $m \geq 0$. A pair of states α,β *communicate with each other* if α has access to β and vice versa. A set of states A has *direct access* to a set of states B if $\sum_{\beta\in B} \tilde{P}^{\xi}_{\alpha\beta} > 0$, for some $\alpha \in A$ and $\xi \in \tilde{K}(\alpha)$. A set of states A is called *communicating* if every pair of states in A communicate.

The decomposition described below produces the partition: $\tilde{\Omega} = \cup_{m=1}^{r} \cup_{p=1}^{n(m)} C_{(m,p)}$ where m is called the *level*, and p is called the *index* of *class* $C_{(m,p)}$. The classes $C_{(m,p)}$ are non-empty and form a partition of $\tilde{\Omega}$. The m-th *level set* denotes the set of $n(m) \geq 1$ classes $\cup_{p=1}^{n(m)} C_{(m,p)}$ having level m. This decomposition has the following properties

(6.1) each class $C_{(m,p)}$ is a communicating set of states

(6.2) for $m \geq 2$, each state in each $C_{(m,p)}$ has access only to $C_{(m,p)}$ and states in level sets $1,2,\dots,m-1$.

(6.3) If $m = 1$, each class $C_{(1,p)}$ is closed under all policies. If $m \geq 2$, each class $C_{(m,p)}$ has direct access to at least one class in the m-1-th level set

(6.4) each state in E constitutes a separate class in the first level set

(6.5) each of the subchains of each of the policies $\varphi \in \Phi$ (cf. lemma 5.2) is contained within one of the classes $C_{(m,p)}$.

The following *procedure* generates this decomposition of $\tilde{\Omega}$:

step 0: Initially, set $\Delta = \tilde{\Omega}$, and define $\bar{\Delta} = \tilde{\Omega} \setminus \Delta$. Initially all level sets are empty

step 1: Let $L = \{\alpha \in \Delta \mid$ for all $\beta \in \Delta$, α has access to β, implies β has access to $\alpha\}$. Decompose L into various communicating sets of states, by using the equivalence classes generated by the relation by communicating on L. Assign these communicating classes to level sets as follows:

if a class has no access to a level set, assign it to level set 1 and increment $n(1)$; otherwise assign it to the m+1-st level set, and increment n(m+1), where the m-th level set is the highest one this class has *direct access* to.

Transfer all states in L from Δ to $\bar{\Delta}$; go to step 3 if Δ is empty, and step 2 otherwise.

step 2: For each state $\alpha \in \Delta$ with $\sum_{\beta\in\bar{\Delta}} \tilde{P}^{\xi}_{\alpha\beta} = 1$ for all $\xi \in \tilde{K}(\alpha)$, add α as a separate class to the m+1-st level set, and increment n(m+1), where the m-th level set is the highest one state α has (direct) access to. Transfer α from Δ to $\bar{\Delta}$. Keep repeating step 2 (with the new Δ and $\bar{\Delta}$) until no more α's can be found. Then go to step 3 if Δ is empty, and go back to step 1 otherwise.

step 3: Let the r^*-th level set $(1 \le r^* \le n^*)$ be the highest level set which is non-empty, and stop the procedure.

The proof that this algorithm is finite uses the bounded number of times step 2 may be performed in succession, and the property that set L is non-empty each time step 1 is executed, so that at least one state is transferred from Δ to $\bar{\Delta}$ during each execution of step 1. This property can be verified for the first execution of step 1, where L consists of the set of subchains of the (stochastic) matrix $[1/\tilde{K}(\alpha) \sum_{\xi\in\tilde{K}(\alpha)} \tilde{P}^{\xi}_{\alpha\beta}]$ because any finite Markov chain has at least one recurrent state. The property holds for subsequent executions of step 1, because termination of step 2 implies that each $\alpha \in \Delta$ has direct access to some state in Δ. By treating the states on Δ as a Markov Chain with positive transition probabilites only where direct access exists, the proof for the first execution of step 1 remains valid.

We next prove the properties (6-1),...,(6-5). Observe that (6-1),...,(6-3) are immediate from the description of the decomposition procedure. Since each state in E has only access to itself, it belongs to the set L which is generated at the first execution of step 1, and constitutes an equivalence class on its own, which proves (6-4). Finally under the assumption that there exists some subchain $\tilde{C}$ for some $\tilde{P}(\varphi)$, $\varphi \in \Phi$ which intersects say $C_{(m,p)}$ and $C_{(n,q)}$ with $(m,p)\neq(n,q)$, the classes would have access to each other, thus contradicting (6-2) and (6-3). This proves (6-5) by contradiction.

For each class $C_{(m,p)}$ $(1 \le m \le r^*;\ 1 \le p \le n(m))$, let

$I(m,p) = \{(n,q)\neq(m,p) \mid C_{(m,p)}$ has direct access to $C_{(n,q)}\}$

$\tilde{\Omega}(m,p) = \cup_{(n,q)\in I(m,p)} C_{(n,q)}$.

Finally, let $\tilde{\Omega}_m = \cup_{\ell=1}^{m} \cup_{p=1}^{n(\ell)} C_{(\ell,p)}$ denote the union of the first m level sets ($1 \le m \le r^*$), with $\tilde{\Omega}_0 = \emptyset$.

THEOREM 6.1. *Fix* $v \in V$.

(a) *The inequalities* (5.2) *which describe the parameter set* $Y(v)$ *decouple as follows:*

(6.6) $$y_\alpha \ge [\tilde{q}_\alpha^\xi + \textstyle\sum_{\beta \in \tilde{\Omega}(m,p)} \tilde{P}_{\alpha\beta}^\xi y_\beta] + \sum_{\beta \in C_{(m,p)}} \tilde{P}_{\alpha\beta}^\xi y_\beta,$$

for all $\alpha \in C_{(m,p)}$; $m = 1,\dots,r^*$; $p = 1,\dots,n(m)$.

(b) *Fix* $m = 1,\dots,r^*$; *let* G *be any (possibly empty) subcollection of classes in the* m+1-st *level set* ($m \ge 0$):
Let $(y_\alpha \mid \alpha \in G \cup \tilde{\Omega}_m)$ *satisfy the inequalities* (6.6). *Then* $(y_\alpha \mid \alpha \in \tilde{\Omega} - G - \tilde{\Omega}_m)$ *may be found such that* $y \in Y(v)$. *In particular, let* $(y_\alpha \mid \alpha \in E)$ *be arbitrary: then* $(y_\alpha \mid \alpha \in F)$ *may be found such that* $y \in Y(v)$.

(c) *(cf. proposition* 6.2*) Fix* $y \in Y(v)$, *and let* $\{\lambda_{m,p} \mid m = 1,\dots,r^*;\ p = 1,\dots,n(m)\}$ *be a set of scalars such that*

(6.7) $$\lambda_{m,p} \ge \max_{(n,q)\in I(m,p)} \lambda_{n,q} \quad \text{for all } m = 1,\dots,r^*;\ p = 1,\dots,n(m)$$

(e.g. $\lambda_{m,p} \ge \lambda_{m-1,q}$ *for all* $m = 2,\dots,r^*$; $p = 1,\dots,n(m)$; $q = 1,\dots,n(m-1)$).

Then $\bar{y} \in Y(v)$, *where* $\bar{y}_\alpha = y_\alpha + \lambda_{m,p}$ *for all* $\alpha \in C_{(m,p)}$; $m = 1,\dots,r^*$; $p = 1,\dots,n(m)$.

(d) $Y(v)$ *is a closed, unbounded, convex polyhedral set containing* $y = 0$ (*i.e.* $\lambda y \in Y(v)$ *for* $0 \le \lambda \le 1$, *if* $y \in Y(v)$).
If $y \in Y(v)$, *then* $[y_\alpha + c_1 \underline{1} + c_2 g^{*\alpha}] \in Y(v)$ *for all scalars* c_1, c_2.

(e) *There exists a bound* $M = M(v)$ *such that for all* $m = 1,\dots,r^*$ *and* $p = 1,\dots,n(m)$:

(6.8) $$\max_{\alpha,\beta \in C_{(m,p)}} |y_\alpha - y_\beta| \le M, \text{ for any } y \in Y(v)$$

In particular, let C *be a subchain of some policy* $\varphi \in \Phi$. *Then* $\max_{\alpha,\beta\in C} \mid y_\alpha - y_\beta \mid \le M$, *for any* $y \in Y(v)$.

PROOF.

(a) (6.6) is immediate from the definition of $\tilde{\Omega}(m,p)$; note that the inequalities are triangularly decomposed since $\tilde{\Omega}(m,p)$ is contained within the first m-1 level sets for $m \ge 2$ and is empty for $m = 1$ (cf. (6.2)).

(b) Choose $y_\alpha = \max_{\beta\in G\cup\tilde{\Omega}_m} y_\beta$, for all $\alpha \in \tilde{\Omega} \setminus G \setminus \tilde{\Omega}_m$ and use the fact that $\tilde{q}_\alpha^\xi \le 0$, for all $\alpha \in \tilde{\Omega}$, $\xi \in \tilde{K}(\alpha)$. The last assertion follows from (6.4) by taking $G = E$ and $m = 0$.

(c) Immediate from (6.6) and the definition of $I(m,p)$.

(d) The fact that $Y(v)$ is a closed, convex, polyhedral set is immediate from (5.2) or (6.6); $\underline{0} \in Y(v)$ follows from $\tilde{q}_\alpha^\xi \le 0$, $\alpha \in \tilde{\Omega}$, $\xi \in \tilde{K}(\alpha)$. The last assertion follows from (5.2), the fact that $g^* = \Pi(f)g^*$ for all $f \in X_i\ L(i)$ and $\tilde{q}_\alpha^\xi \le 0$, for all $\alpha \in \tilde{\Omega}$, $\xi \in \tilde{K}(\alpha)$. This in turn exhibits the unboundedness of $Y(v)$.

(e) We prove (6.8) for a fixed class $C = C_{(m,p)}$ ($m = 1,\dots,r^*$; $p = 1,\dots,n(m)$); part (e) then follows from the fact that the number of classes $C_{(m,p)}$ is finite. First note that either C was generated in step 2 of the above decomposition procedure in which case it contains a single element and (6.8) is trivially met with $M = 0$, or C is a closed set of mutually communicating states. In the latter case there exists a (possibly randomized) policy φ which has C as one of its subchains.
Next, rewrite (5.4) as follows, by introducing a slack vector $t \ge 0$:

(6.9) $$y = \tilde{q}(\varphi) + t + \tilde{P}(\varphi)y.$$

Let $\{\tilde{\pi}^C(\varphi)_\alpha \mid \alpha \in C\}$ denote the unique equilibrium distribution of $\tilde{P}(\varphi)$ on C. Multiply (6.9) with $\tilde{Z}(\varphi)$. Then, since $\tilde{Z}(\varphi)_{\beta\gamma} = 0$, for $\beta \in C$, $\gamma \notin C$ (cf. (2.3)):

$$y_\beta = \sum_{\gamma\in C} \tilde{Z}(\varphi)_{\beta\gamma}\ (\tilde{q}(\varphi)_\gamma + t_\gamma) + \sum_{\gamma\in C} \tilde{\pi}^C(\varphi)_\gamma y_\gamma, \qquad \text{all } \beta \in C.$$

Part (e) follows with the choice $M = 2 \max_{\beta\in C}\left\{\sum_{\alpha\in C} \mid \tilde{Z}(\varphi)_{\beta\alpha}(\tilde{q}(\varphi)_\alpha + t_\alpha) \mid\right\}$ provided one shows that $[t_\alpha \mid \alpha \in C]$ are bounded uniformly in y. However, by multiplying (6.9) with $\tilde{\pi}^C(\varphi)$ one obtains:

$$- \sum_{\beta\in C} \tilde{\pi}^C(\varphi)_\beta \, \tilde{q}(\varphi)_\beta = \sum_{\beta\in C} \tilde{\pi}^C(\varphi)_\beta \, t_\beta .$$

The boundedness of $[t_\beta \mid \beta \in C]$ follows since $\tilde{\pi}^C(\varphi)_\beta > 0$ for $\beta \in C$. The last assertion follows from (6.4). □

A *ray* for the solution set to a set of linear inequalities is a solution to the corresponding homogeneous set of inequalities (cf. [26]). The rays to Y(v) are therefore the solutions $u = (u_1,\ldots,u_{n^*})$ to:

$$u_\alpha \geq \sum_{\beta=1}^{n^*} \tilde{P}^\xi_{\alpha\beta} \, u_\beta, \qquad \alpha \in F,\ \xi \in \tilde{K}(\alpha).$$

Define U as the set of rays to Y(v) and remark that U is independent of $v \in V$, since F, $\tilde{K}(\alpha)$ and $\tilde{P}^\xi_{\alpha\beta}$ are. U has the following important and easily verified properties:

(a) if $u, \hat{u} \in U$, then $c_1 u + c_2 \hat{u} \in U$ for all $c_1, c_2 \geq 0$
(b) if $v \in V$, $y \in Y(v)$ and $u \in U$, then $y + c\,u \in Y(v)$, for all $c \geq 0$.

Observe that th. 6.1 parts (b), (c) and (d) apply to the set U as well as to Y(v). Note from the proof of part (d) of th. 6.1 that the vectors u with $u_\alpha = c\, g^{*\alpha}$ are members of U, for any scalar c. Likewise, it follows from the proof of part (c) that

$$\{u \mid u_\alpha = \lambda_{m,p} \text{ for all } \alpha \in C_{(m,p)};\ m = 1,\ldots,r^*;\ p = 1,\ldots,n(m)\} \subseteq U$$

provided that the scalars $\lambda_{m,p}$ satisfy (6.7). The following proposition, finally, gives an additional characterization of the set U and the vector g^*:

<u>PROPOSITION 6-2:</u>

(a) *If* $u \in U$, *then* $u_\alpha = u_\beta$ *for all* $\alpha, \beta \in C_{(m,p)}$; $m = 1,\ldots,r^*$; $p = 1,\ldots,n(m)$.
(b) *There are* $\sum_{m=1}^{r^*} n(m) \leq n^*$ *constants* $\{x_{m,p} \mid m = 1,\ldots,r^*;\ p = 1,\ldots,n(m)\}$ *satisfying*

(6.10) $u_\alpha = x_{m,p}$ for all $\alpha \in C_{(m,p)}$

(6.11) $x_{m,p} \geq \sum_{(n,q)\in I(m,p)} \rho^{\xi}_{\alpha;n,q} \, x_{n,q}$

for all $m = 1,\ldots,r^*$; $p = 1,\ldots,n(m)$; $\alpha \in C_{(m,p)}$ and $\xi \in \tilde{K}(\alpha)$ with $\sum_{\beta\in C(m,p)} \tilde{P}^{\xi}_{\alpha\beta} < 1$ where

$$\rho^{\xi}_{\alpha;n,q} = [1 - \sum_{\beta\in C_{(m,p)}} \tilde{P}^{\xi}_{\alpha\beta}]^{-1} \sum_{\beta\in C_{(n,q)}} \tilde{P}^{\xi}_{\alpha\beta}$$

i.e. U is a closed convex, unbounded, polyhedral set containing $u = 0$ with dimension $\sum_{m=1}^{r^*} n(m)$, which is described by the *completely decoupled* set of (in)equalities (6.10) and (6.11)

(c) The components of $[g^{*\alpha} \mid \alpha \in \tilde{\Omega}]$ satisfy:

$$g^{*\alpha} = g^*_{m,p} \text{ for all } \alpha \in C_{(m,p)}, \text{ where}$$

$$g^*_{m,p} = \sum_{(n,q)\in I(m,p)} \rho^{\xi}_{\alpha;n,q} \, g^*_{n,q}$$

for all $m = 1,\ldots r^*$; $p = 1,\ldots,n(m)$; $\alpha \in C_{(m,p)}$ and $\xi \in \tilde{K}(\alpha)$ with $\sum_{\beta\in C_{(m,p)}} \tilde{P}^{\xi}_{\alpha\beta} < 1$.

PROOF.

(a) Immediate from the proof of part (e) of th. 6.1.

(b) (6.10) follows from part (a) and (6.11) follows from (6.10) and (6.6)

(c) part (c) follows from the fact that $[c\, g^{*\alpha} \mid \alpha \in \tilde{\Omega}] \in U$ for all (positive and negative) scalars c, using (6.10) and (6.11). □

Example 2 below has $N = 7$, $g^*_i = 0$, $L(i) = K(i)$ for all i:

$$R^* = \cup_{i=1}^{7} R^{*i} \text{ with } R^{*i} = \{i\}, \text{ i.e. } n^* = 7.$$

$$E = \{\alpha = 1\};\ F = \{\alpha = 2,3,\ldots,7\};$$

$$C_{(1,1)} = \{1\};\ C_{(1,2)} = \{2,3\};\ C_{(2,1)} = \{4\};\ C_{(2,2)} = \{5,6\};$$

$$C_{(2,3)} = \{7\}$$

V is the solution set to the following decoupled set of inequalities:

$$\alpha = 1: \quad v_1 \text{ arbitrary}$$

$$\alpha = 2,3: \quad q_2^2 \le v_2 - v_3 \le -q_3^2$$

$$\alpha = 4: \quad v_4 \ge q_4^2 + v_1$$

$$\alpha = 5,6: \quad q_5^2 \le v_5 - v_6 \le -q_6^2$$

$$v_6 \ge q_6^3 + .5v_1 + .5v_3$$

$$\alpha = 7: \quad v_7 \ge q_7^2 + .5(v_1 + v_2).$$

<u>Example 2</u>

i	k	q_i^k	P_{i1}^k	P_{i2}^k	P_{i3}^k	P_{i4}^k	P_{i5}^k	P_{i6}^k	P_{i7}^k
1	1	0	1						
2	1	0		1	0				
	2	$q_2^2 < 0$		0	1				
3	1	0		0	1				
	2	$q_3^2 < 0$		1	0				
4	1	0				1			
	2	$q_4^2 < 0$	1						
5	1	0					1	0	
	2	$q_5^2 < 0$					0	1	
6	1	0					0	1	
	2	$q_6^2 < 0$					1	0	
	3		.5	.5					
7	1	0							1
	2	$q_7^2 < 0$	.5	.5					

Absent P_{ij}^k are zero.

VII. ACKNOWLEDGMENT

We wish to express our sincere thanks to Dr. Henk Tijms, for his useful comments and careful reading of this and previous versions of this paper.

REFERENCES

[1] BATHER, J., *Optimal decision procedures for finite Markov Chains,* Part III, Adv. Appl. Prob. 5 (1973), 541-554.

[2] BELLMAN, R., *A Markovian Decision Process,* J. Math. Mech. 6 (1957), 679-684.

[3] BELLMAN, R., *Functional Equations in the Theory of Dynamic Programming,* V. *Positivity and Quasi-Linearity,* Proc. Nat. Acad. Sci. U.S.A. 41 (1955), 743-746.

[4] BLACKWELL, D., *Discrete Dynamic Programming,* Ann. Math. Statistics 33 (1962), 719-726.

[5] BROWN, B., *On the iterative method of dynamic programming on a finite state space discrete time Markov Process,* Ann. Math. Statist. 36 (1965), 1279-1285.

[6] DeCANI, J., *A Dynamic Programming Algorithm for Embedded Markov Chains when the Planning Horizon is at Infinity,* Management Sci. 10 (1964), 716-733.

[7] DENARDO, E., *Markov Renewal Programs with small interest rates,* Ann. of Math. Statistics 42 (1971), 447-496.

[8] DENARDO, E. & B. FOX, *Multichain Markov Renewal Programs,* SIAM, J. Appl. Math. 16 (1968), 468-487.

[9] DERMAN, C., *Finite State Markovian Decision Processes,* Academic Press, New York (1970).

[10] GALPERIN, A., *The General Solution of a Finite System of Linear Inequalities*, Math. of O.R. 1 (1976), 185-196.

[11] HOWARD, R., *Dynamic Programming and Markov Processes*, John Wiley, New York (1960).

[12] HOWARD, R., *Semi Markovian Decision Processes,* Bult. Int. Stat. Inst. 40 (1963), 625-652.

[13] JEWELL, W., *Markov Renewal Programming*, Oper. Res. 11 (1963), 938-971.

[14] KEMENY, J. & J. SNELL, *Finite Markov Chains*, Van Nostrand, Princeton (1961).

[15] LANERY, E., *Etude Asymptotique des Systèmes Markoviens à Commande*, R.I.R.O. 1 (1967), 3-56.

[16] LANERY, E., *Complements à l'étude Asymptotique des Systèmes Markoviens à Commande*, I.R.I.A., Rocquencourt, France (1968).

[17] LEMBERSKY, M., *Preferred Rules in continuous time Markov Decision Processes*, Man. Sci. 21 (1974), 348-357.

[18] MILLER, B., *Finite state continuous time Markov Decision Processes with an infinite planning horizon*, J.M.A.A. 22 (1968), 552-569.

[19] ROMANOVSKII, I.V., *The Turnpike Theorem for Semi-Markov Decision Processes*, in: LINNIK, Yu.V., *Theoretical Problems in Math Statistics*, American Mathematical Society, Providence (1972), 249-267, translated from the Proceedings of the Steklov Institute of Mathematics 111 (1970).

[20] ROMANOVSKY, I., *On the solvability of Bellman's functional equation for a Markovian Decision Process*, J. of Math. Anal. and Appl. 42 (1973), 485-498.

[21] SCHWEITZER, P.J., *Perturbation theory and Markovian Decision Processes*, Ph.D. dissertation, MIT (1965) MIT Operations Research Center Report 15.

[22] SCHWEITZER, P.J., *Perturbation theory and undiscounted Markov Renewal Programming*, Oper. Res. 17 (1969), 716-727.

[23] SCHWEITZER, P.J., *Perturbation Theory and Finite Markov Chains*, J. Applied Probability 5 (1968), 401-413.

[24] SCHWEITZER, P.J., & A. FEDERGRUEN, *Asymptotic Value Iteration for Undiscounted Markov Decision Problems*, Math. Center Report BW 44/76 (1976).

[25] SCHWEITZER, P.J. & A. FEDERGRUEN, *Foolproof Convergence in multichain policy iteration*, I.B.M. Thomas J. Watson Research Center report RC 5894 (1976), (to appear in J.M.A.A.).

[26] WILLIAMS, A., *Complementary Theorems for linear programming*, SIAM Review 12 (1970), 135-137.