stichting
mathematisch
centrum

MC

DEPARTMENT OF OPERATIONS RESEARCH BW 60/76 FEBRUARY

P.J. SCHWEITZER & A. FEDERGRUEN

THE FUNCTIONAL EQUATIONS OF UNDISCOUNTED MARKOV RENEWAL PROGRAMMING

Prepublication

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), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

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 a.o. in the Undiscounted Markov Renewal Programming. We show that the solution set is a connected, though non-convex set whose members are unique up to n^* constants, characterize n^* and show that these n^* degrees of freedom are locally rather than globally independent.

Our results generalize those obtained in ROMANOVSKY [15] 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 paper is not for review; it is meant for publication elsewhere.

I. INTRODUCTION

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

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

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

where

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

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 P 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 [1,2], BLACKWELL [3], HOWARD [9,10], DE CANI [5], JEWELL [11], DENARDO & FOX [7], DENARDO [6], DERMAN [8], SCHWEITZER [16,17,18]).

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 non-convex set whose members are unique up to n^* constants, characterize n^*, and show that these n^* degrees of freedom are locally rather than globally independent.

Our results generalize those obtained in ROMANOVSKY [15] 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 some remarks are made with respect to a triangular decomposition of the set V.

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. [3],[12]):

$$\Pi(f) = P(f)\Pi(f) = \Pi(f)P(f) = \Pi(f)^2 = \Pi(f)Z(f) = Z(f)\Pi(f) \tag{2.1}$$

$$[I - P(f)]Z(f) = Z(f)[I - P(f)] = I - \Pi(f) \tag{2.2}$$

$$Z(f) = I + \lim_{a \uparrow 1} \sum_{n=0}^{\infty} a^n [P(f)^n - \Pi(f)]. \tag{2.3}$$

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

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

where $\pi^m(f)$ 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 absorbtion in $C^m(f)$, starting from state i (cf. [6] and [18]). Observe $\sum_i \pi_i^m(f) = 1$ and $\pi^m(f)P(f) = \pi^m(f)$.

Let $R(f) = \{j \mid \Pi(f)_{jj} > 0\}$, i.e. R(f) is the set of recurrent states for P(f). Note that $\phi^m(f) = P(f)\phi^m(f)$ for all m and that the vecotrs $\phi^m(f)$ 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) $$x = \sum_{m=1}^{n(f)} a_m\phi^m(f)$$

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

LEMMA 2.1. *Fix* $f \in S_R$, *and let the vector* b *satisfy* $\Pi(f)b = 0$. *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 $[I - P(f)]x \ge b$ by $\Pi(f) \ge 0$, yields $0 = \Pi(f)[I - P(f)]x \ge \Pi(f)b = 0$, 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 *such that* (a) $h_{ik} = 1$, (b) *for every* (j,r) $h_{jr} = 1$ *only if* $f_{jr} > 0$, (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. Since $C^m(f)$ is closed for P(f), it is closed for any h meeting (b). Now, let $h_{ik} = 1$. 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 some 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 P holds.

P: 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 P holds for all pure policies, it holds for all randomized policies.

<u>LEMMA 2.3</u>. (*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) $$g_i = g(f)_i = \sum_{m=1}^{n(f)} \phi^m_i(f) g^m(f),$$

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 P, it follows from the proof of lemma 1 of [6] 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) = {1,2}; $H^k_{ij} = \delta_{ij}$ for all i,j,k.

i	k	p^k_{i1}	p^k_{i2}	p^k_{i3}	p^k_{i4}	q^k_i
1	1	0	1	0	0	0
2	1	1	0	0	0	0
3	1	1	0	0	0	$q^1_3\le 0$
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 [17] 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 contant a.

In addition, we observe that the Policy Interation Algorithm (PIA) (cf. [5], [7], [11]) is not guaranteed to converge, if unwise choices for the additive constants in (2.8) are made. Consider the above example with $q_3^1 < 0$, $f^{1\#} = (1,1,2,1)$ and $f^{2\#} = (1,1,2,2)$. Then $v(f^1) = \lambda[1\ 1\ 0\ .8] + \mu[0\ 0\ 1\ .2]$ and $v(f^2) = \nu[1\ 1\ 0\ 1] + \rho[0\ 0\ 1\ 0]$, for arbitrary λ,μ,ν,ρ. Choosing $q_3^1 + \lambda \leq \mu < \lambda$ and $\rho > \nu$, f^1 and f^2 follow each other in the PIA. Fortunately, PIA cycling can be prevented by preserving the old additive constant in a subchain, whenever the subchain is preserved (see also [20]).

III. PROPERTIES OF MAXIMAL GAIN POLICIES

We first introduce some notations. Define the *maximal gain rate*

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

For any $v \in V$, define

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

and

$$b(v,f)_i = \sum_{k \in K(i)} b(v)_i^k = [q(f) - H(f)g^* + P(f)v - v]_i$$

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 [8] that there exists a pure policy that attains the N suprema in (3.2) simultaneously. Hence $g_i^* = \max_{f \in S_P} g(f)_i$.

Accordingly define:

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

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) *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) From the proof of lemma 2.3 we know that 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 [7] 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:

$$\max_{f \in S_{PMG}} Z(f)[q(f) - H(f)g^*]_i = \max_{f \in \bar{S}_{PMG}} Z(f)\{\bar{q}(f) - \bar{g}^*\}_i. \quad \text{for } i = 1,\ldots,N.$$

Use theorem 4 of [3] 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 P, a minor modification of the proof of lemma 4 of [7], 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 [7].

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

(1) From the definition of $b(v)_i^k$, 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 (d).

(2) If $f \in S_{RMG}$, $g^* = P(f)g^*$ follows from part (d). Hence $f_{ik} > 0$ implies $k \in L(i)$ and $b(v)_i^k \leq 0$. So $b(v,f) \leq 0$, for any $v \in V$. Since we know from part (d) 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. $f_{ik} > 0$ implies $b(v)_j^k = 0$. □

Define next

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

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* $1 \le \alpha \le n^*$, $g^*_i = g^{*\alpha}$ *(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* $\{R^{(k)}\}_{k=1}^m$.

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

PROOF.

(a) Fix a state i, 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. Using theorem 3.1. part (e), one verifies that $h \in S_{PMG}$, and $i \in R(h)$. Thus the right-hand side of (a) is included in 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 $\mathcal{C} = \{C^m(f^r) \mid 1 \le r \le M;\ 1 \le m \le n(f^r)\}$:

Let $C \sim C'$, if there exists $\{C^{(1)}=C, C^{(2)},\dots,C^{(n)}=C'\}$ with $C^{(i)} \in \mathcal{C}$ and $C^{(i)} \cap C^{(i+1)} \neq \emptyset$ 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 a same equivalence class communicate with each other. Hence, $R^* = R(f^*)$.

(c) Assume $P(f)$ has a subchain $C^m(f)$ that intersects say R^{*1} and R^{*2}. Then a policy f^{**} with $\{k \mid f^{**}_{ik} > 0\} = \{k \mid f^*_{ik} > 0\}$ and $\{k \mid f^{**}_{ik} > 0\} = \{k \mid f_{ik} > 0\} \cup \{k \mid f^*_{ik} > 0\}$ 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 (c) of theorem 3.1.

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

REMARK 1. Note that as a result of part (f) of the above theorem, the policy f^* that was constructed in the proof of part (b), belongs to S^*_{RMG}. Verify that the definition of f^* implies any subchain of a maximal gain policy to be contained in a subchain of $P(f^*)$.

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 maxima in (1.1) and (1.2)$\} \subseteq S_{PMG}$. Part (a) of theorem 3.2 in combination with part (a) of theorem 3.1 establish $R^* = \{i \mid i \in R(f), f \in S_P(v)\}$. 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 remark 1). Finally, define f^* by $\{k \mid f^*_{ik} > 0\} = L(i)$ for $i \in \Omega - R^*$, and $\{k \mid f^*_{ik} > 0\} = \{k \in L(i) \mid b(v)_i^k = 0, \sum_{j \in R^{*\alpha}} P_{ij}^k = 1\}$ for $i \in R^{*\alpha}$ $(\alpha = 1,\dots,n^*)$.

VI. PROPERTIES OF V

Some basic properties of V are given by:

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

(a) V *is closed an 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 unitary).*

(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.* [2],[16].*)* $v \in V$, *if and only if*

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

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 (2.9), 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) acheives all N maxima in (4.1).

Conversely, if v satisfies (4.1), we define:

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

and show both $\tilde{v} \geq v$ and $\tilde{v} \leq v$, 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 (2.9):

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

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

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

$$(4.3) \qquad v_i \leq \tilde{v}_i = [q(h) - H(h)g^* + P(h)v]_i.$$

Multiply (4.3) with $\Pi(h) \geq 0$ in order to get $0 \leq \Pi(h)[q(h) - H(h)g^*] \leq 0$, the latter inequality following from (2.9) and $g(h) \leq g^*$. Hence (4.3) is an equality for $i \in R(h)$, and so $h \in S_{PMG}$, by part (e) of theorem 3.1.

Using lemma 2.1, (4.3) implies $v \le Z(h)[q(h) - H(h)g^*] + \Pi(h)v$. Insert 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 $h \in S_{PMG}$ achieves the N maxima in (4.1), multiply the equality portion of this inequality with $Z(h)^{-1}$ to show that it achieves the N maxima in (1.2), as well as the N maxima in (1.1), since $h_{ik} = 1$ 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

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

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

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

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

<u>THEOREM 4.2.</u>

(a)

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

Hence, V^R *is a closed, convex 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 $i \in R(h)$. 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.

(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: $v(\lambda)_i = \lambda v_i + (1-\lambda)\bar{v}_i$, $i \in R^*$ and $v(\lambda)_i = \max_{f\in S_{PMG}} \{Z(f)[q(f) - H(f)g^*]_i + \sum_{j\in R^*} \Pi(f)_{ij}v(\lambda)_j\}$ 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.

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

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 are equivalent to the linear (in)equalities $b(v)_i^{k(i)} = 0$ for $i = 1,\dots,N$ and $b(v)_i^k \leq 0$ 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 2. (4.8), hence convexity of V is trivially met if either (1) $R^* = \Omega$, (2) $L(i)$ is a singleton for each $i \in \Omega - R^*$, or (3) there is only one maximal gain policy.

In addition $n^* = 1$ is sufficient for the convexity of V as well. This follows by considering a $f^* \in S^*_{RMG}$. By theorem 4.2 part (b), we obtain that for each $v \in V$, there exists a relative value vector $v(f^*)$ such that $v_i = v(f^*)_i$, $i \in R^*$. $P(f^*)$ being unichained, it follows that $v(f^*)$ is unique up to a multiple of $\underline{1}$, hence $(v_i | i \in R^*)$ is unique up to an additive constant. Using (4.5), we conclude that $v \in V$ is unique up to a multiple of $\underline{1}$.

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

$$(4.9) \qquad v(n+1)_i = \max_{k \in K(i)} \{q^k_i + \sum_j P^k_{ij} v(n)j\},$$

with $v(0)$ a given vector.

It is well known that $\{v(n) - ng^*\}^\infty_{n=1}$ may fail to converge. In a forthcoming paper [19] it will be shown that there exists an integer J such that

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

exists for all i, with $u^{(r+J)}_i = u^{(r)}_i$ (previous proofs in [4] and [13] are both incorrect).

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

The relation between 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* $v \le \bar{v}$ *with equality for components in* R^*.

PROOF. Note that for all $i \in \Omega$: $u^{(r+1)}_i = \max_{k \in K(i)} \{q^k_i - g^*_i + \sum_j P^k_{ij} u^{(r)}_j\}$, since for all n sufficiently large the maximizing alternatives in (4.9) belong to L(i) as observed in [4] and [13].

Since $v = \frac{1}{J}\sum_{r=0}^{J-1} u^{(r)}$, 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) $v_i = \max_{f \in S_{PMG}} \{Z(f)[q(f) - g^*]_i + \sum_{j \in R^*} \Pi(f)_{ij} v_j\}$ 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 $(y_1,\ldots,y_{n^*})$.

ROMANOVSKY [15] 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 $v, \tilde{v} \in V$ have the property: $\tilde{v}_i - v_i = \text{constant} = y_\alpha$ for $i \in R^{*\alpha}$, $\alpha = 1,\ldots,n^*$.

By fixing $v^o \in V$ and picking these n^* constants, one thus determines $(\tilde{v}_i | i \in R^*)$ 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 v.

<u>THEOREM 5.1</u>. *Let* $v \in V$. *The following are equivalent:*

(a) $v + x \in V$

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

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

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

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

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

$$\alpha = 1,\dots,n^*;\ i \in R^{*\alpha},\ f \in S_{PMG}.$$

PROOF.

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

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

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

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

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

The following theorem is obvious from the definition of $Y(v)$, theorem 4.1 part (a) and the fact that:

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

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

THEOREM 5.2. *For any* $v \in V$, $Y(v)$ *is a closed, convex polyhedral set containing* $y = 0$, *(i.e.* $\lambda y \in V$, *for* $\lambda \in [0,1]$ *if* $y \in Y(v)$*).*

Furthermore, $Y(v)$ *is unbounded as* $[y_\alpha] \in Y(v)$, *implies* $[y_\alpha + c_1 + c_2 g^{*\alpha}] \in Y(v)$, *for any scalars* c_1, c_2.

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:

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

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

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

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

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

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

The following theorem expresses that $(y_{\alpha} | \alpha\in E)$ are fully independent degrees of freedom:

<u>THEOREM 5.3</u>.

(a) *Let* $(y_{\alpha}|\alpha\in E)$ *be arbitrary. Then* $(y_{\alpha}|\alpha\in F)$ *can be found such that* $y \in Y(v)$.

(b) *If* $y \in Y(v)$, *then after arbitrary decreases in any of the* y_{α}, $\alpha \in E$, y *is still in* $Y(v)$.

<u>PROOF</u>.

(a) Take $y_{\alpha} = \max_{\beta\in E} y_{\beta}$, $\alpha \in F$.

(b) The inequalities (5.4) are either unaffected or strengthened by decreasing $(y_{\alpha}|\alpha\in E)$. □

A *ray* for the solution set to a set of linear inequalities is a solution to the corresponding homogeneous set of inequalities (cf. [22]). The rays to $Y(v)$ are therefore the solutions $(y_1,\ldots,y_{n^*})$ to:

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

Define U as the set of rays to Y(v) and remark that U is independent of v, since F, $\tilde{K}(\alpha)$, $\tilde{P}^{\xi}_{\alpha\beta}$ are. Since U is the set of rays to Y(v), it 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 $y + cu \in Y(v)$ for all $c \geq 0$

REMARK 3. Theorem 5.3 applies to U as well as to Y(v).

Note from theorem 5.2 and theorem 5.3 that the vectors $\bar{u}$ with $\bar{u}_\alpha = cg^{*\alpha}$ and $\bar{\bar{u}}$, with $\bar{\bar{u}}_\alpha = c$, $\alpha \in F$ and $\bar{\bar{u}}_\alpha \leq c$, $\alpha \in E$ are members of U, for any scalar c. Additional properties of U are discussed in theorem 5.4 and section 6.

In order to show that Y(v) is an n^*-dimensional polyhedral set, we need the following discrete time Markovian model with state space $\{1,\ldots,n^*\}$:
For $\alpha \in F$, let $\tilde{K}(\alpha)$ be the set of feasible decision. For $\xi \in \tilde{K}(\alpha)$, let $\tilde{q}^{\xi}_{\alpha}$ and $\tilde{P}^{\xi}_{\alpha\beta}$ 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 $\phi \in \Phi$, the quantities $\tilde{q}(\phi)$, $\tilde{P}(\phi)$, $\tilde{\Pi}(\phi)$ and $\tilde{Z}(\phi)$ 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 $\tilde{q}(\phi) \leq 0$ for any $\phi \in \Phi$. The following theorem characterizes the subchains of $\tilde{P}(\phi)$ on F:

THEOREM 5.4. *(Properties of subchains of $\tilde{P}(\phi)$ on F).*
Fix $v \in V$. Suppose for some policy $\phi \in \Phi$, $\tilde{P}(\phi)$ has a subchain $C \subseteq F$. Then

(a) *C has at least two members.*

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

(c) *There exists a bound $M = M(v)$ such that*

$$\max_{\alpha,\beta \in C} |y_\alpha - y_\beta| \leq M \qquad \text{for any } y \in Y(v).$$

(d) *If $\bar{y}$ is a ray to Y(v) then $\bar{y}_\alpha = \bar{y}_\beta$, for all $\alpha,\beta \in C$.*

<u>PROOF</u>.

(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:

(5.6) $\qquad \alpha \in C,\ i \in S(\alpha)$ 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 $f(\alpha) \in S_{PMG}$, $b(v,f(\alpha)) \le 0$ with equality for components in $R(f(\alpha))$. Hence, using (2.3), $0 = \tilde{q}(\phi)_\alpha =$
$= \sum_{j \notin R(f(\alpha))} b(v,f(\alpha))_j = \sum_{j \notin R(f(\alpha))} \sum_{n=0}^{\infty} [P(f(\alpha))]^n_{i(\alpha)j} \cdot b(v,f(\alpha))_j$.
Hence:

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

We now exhibit a policy $f^o \in S_{RMG}$ with the contradictory 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 (5.7) 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 $i \in R^o$, (5.6) 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 exist a $j \notin R^{*\alpha}$, and an integer $n \ge 1$, 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.

(c) Introduce a slack vector $t \geq 0$ and rewrite (5.4) as:

$$y = \tilde{q}(\phi) + t + \tilde{P}(\phi)y. \tag{5.8}$$

Let $\{\tilde{\pi}^C(\phi)_\alpha \mid \alpha \in C\}$ denote the unique equilibrium distribution of $\tilde{P}(\phi)$ on C. Multiply (5.8) with $\tilde{Z}(\phi)$. Then, since $\tilde{Z}(\phi)_{\beta\gamma} = 0$ for $\beta \in C$, $\gamma \notin C$ (cf. (2.3)):

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

Part (c) follows with the choice $M = 2\max_{\beta \in C}\{\sum_{\alpha \in C} |\tilde{Z}(\phi)_{\beta\alpha}|\,[|\tilde{q}(\phi)_\alpha| + t_\alpha]\}$ provided one shows that $[t_\alpha \mid \alpha \in C]$ are bounded uniformly in y. However, by multiplying (5.7) with $\tilde{\pi}^C(\phi)$ one obtains:

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

The boundedness of $[t_\beta \mid \beta \in C]$ follows since $\tilde{\pi}^C(\phi)_\beta > 0$ for $\beta \in C$.

(d) Use part (c) and (5.5). □

Together part (b) of theorem 5.4 and the choice $\tilde{q}^{\xi_o}_\alpha = -1$, for $\alpha \in E$ imply:

<u>COROLLARY 5.1</u>. $\tilde{g}^*_\alpha < 0$ *for* $\alpha = 1,\ldots,n^*$.

<u>THEOREM 5.5</u>. (*Cf. theorem* 3 *of* [15].) *Fix* $v \in V$. *Given any* $\{y_\alpha \mid \alpha \in E\}$ *there exist* $\{y_\alpha \mid \alpha \in F\}$ *such that*

$$y_\alpha > \tilde{q}^\xi_\alpha + \sum_{\beta=1}^{n^*} \tilde{P}^\xi_{\alpha\beta} y_\beta, \qquad \textit{for all } \alpha \in F,\ \xi \in \tilde{K}(\alpha) \tag{5.9}$$

holds with strict inequality.

<u>PROOF</u>. It suffices to show that there exists a solution y^o to (5.9) for some $\{y^o_\alpha \mid \alpha \in E\}$ since a solution for any $\{y_\alpha \mid \alpha \in E\}$ is then obtained by adding a ray u with $u_\alpha = y_\alpha - y^o_\alpha$, for $\alpha \in E$ (cf. remark 3).

Since $\tilde{q}^{\xi_o}_\alpha = -1$ and $\tilde{P}^{\xi_o}_{\alpha\alpha} = 1$, for $\alpha \in E$, the solution set to (5.9) is not altered by adding the inequalities $y_\alpha > \tilde{q}^{\xi_o}_\alpha + \sum_{\beta=1}^{n^*} \tilde{P}^{\xi_o}_{\alpha\beta} y_\beta$, $\alpha \in E$. Now, assume to the contrary, that the solution set of (5.9) is empty. Then for the LP-problem:

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

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

we have min $Z \geq 0$, which according to theorem 2 of [14], implies $\max_{\alpha=1,\ldots,n^*} \tilde{g}^*_\alpha \geq 0$. This contradicts corollary 5.1. □

Since the solution set to (5.9) is open, for any y satisfying (5.9), 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.

VI. TRIANGULAR DECOMPOSITION OF Y(v)

Define the following partition of F:

$F^\ell = \{\alpha \in F \mid$ for every $\phi \in \Phi$, α reaches E with certainty under $\tilde{P}(\phi)\}$

$F^t = \{\alpha \in F \mid \alpha$ is transient under any $\tilde{P}(\phi)$, $\phi \in \Phi$, but $\alpha \notin F^\ell\}$

$F^r = \{\alpha \in F \mid \alpha$ is recurrent for some $\tilde{P}(\phi)$, $\phi \in \Phi\}$.

Note that $\sum_{\beta \in F^\ell \cup E} \tilde{P}^\xi_{\alpha\beta} = 1$ for $\alpha \in F^\ell$, $\xi \in \tilde{K}(\alpha)$.

The set of inequalities (5.2) then decouples into 3 parts:

$$(6.1)\qquad y_\alpha \geq \left[\tilde{q}^\xi_\alpha + \sum_{\beta\in E\cup(F\setminus F^t)} \tilde{P}^\xi_{\alpha\beta} y_\beta\right] + \sum_{\beta\in F^t} \tilde{P}^\xi_{\alpha\beta} y_\beta, \qquad \alpha \in F^t,\ \xi \in \tilde{K}(\alpha)$$

$$(6.2)\qquad y_\alpha \geq \left[\tilde{q}^\xi_\alpha + \sum_{\beta\in E} \tilde{P}^\xi_{\alpha\beta} y_\beta\right] + \sum_{\beta\in F^\ell} \tilde{P}^\xi_{\alpha\beta} y_\beta, \qquad \alpha \in F^\ell,\ \xi \in \tilde{K}(\alpha)$$

$$(6.3)\qquad y_\alpha \geq \left[\tilde{q}^\xi_\alpha + \sum_{\beta\in E\cup(F\setminus F^r)} \tilde{P}^\xi_{\alpha\beta} y_\beta\right] + \sum_{\beta\in F^r} \tilde{P}^\xi_{\alpha\beta} y_\beta, \qquad \alpha \in F^r,\ \xi \in \tilde{K}(\alpha).$$

The above decomposition implies that the following vectors belong to U: $u_\alpha = c_1$, $\alpha \in E$; $u_\alpha = c_2$, $\alpha \in F^\ell$; $u_\alpha = c_3$, $\alpha \in F^t \cup F^r$; for all c_1, c_2, c_3 with $c_1 \leq c_2 \leq c_3$. For $\phi \in \Phi$, let $W(\phi) = [P(\phi)_{\alpha\beta}]_{\alpha,\beta\in F^\ell\cup F^t}$.

Then $W(\phi)$ is a substochastic transient matrix, with $\lim_{n\to\infty} W(\phi)^n = 0$ and $[I - W(\phi)]^{-1} = \sum_{n=0}^\infty W(\phi)^n$ exists and is non-negative. Then, taking together (6.1) and (6.2) and using the proof of lemma 1 of [7], we obtain:

$$(6.4)\qquad y_\alpha \geq \max_{\phi\in\Phi} \sum_{\beta\in F^\ell\cup F^t} [I - W(\phi)]^{-1}_{\alpha\beta} [\tilde{q}(\phi)_\beta + \sum_{\gamma\in E\cup F^r} \tilde{P}(\phi)_{\beta\gamma} y_\gamma], \qquad \alpha \in F^t \cup F^\ell.$$

Insert (6.4) into (6.3) in order to obtain:

$$(6.5)\qquad y_\alpha \geq \hat{q}^{\xi,\phi}_\alpha + \sum_{\beta\in E\cup F^r} \hat{P}^{\xi,\phi}_{\alpha\beta}\, y_\beta, \qquad \text{all } \alpha \in F^r,\ \xi \in \tilde{K}(\alpha),\ \phi \in \Phi,$$

where

$$\hat{q}^{\xi,\phi}_\alpha = \tilde{q}^\xi_\alpha + \sum_{\beta\in F^\ell\cup F^t} \tilde{P}^\xi_{\alpha\beta} \sum_{\gamma\in F^\ell\cup F^t} [I - W(\phi)]^{-1}_{\beta\gamma}\, \tilde{q}(\phi)_\gamma$$

$$\hat{P}^{\xi,\phi}_{\alpha\beta} = \tilde{P}^\xi_{\alpha\beta} + \sum_{\gamma\in F^\ell\cup F^t} \tilde{P}^\xi_{\alpha\gamma} \sum_{\delta\in F^\ell\cup F^t} [I - W(\phi)]^{-1}_{\gamma\delta}\, \tilde{P}(\phi)_{\delta\beta}.$$

Notice that $\hat{q}^{\xi,\phi}_\alpha \leq 0$, and $\hat{P}^{\xi,\phi}_{\alpha\beta} \geq 0$ with $\sum_{\beta\in E\cup F^r} \hat{P}^{\xi,\phi}_{\alpha\beta} = 1$.

Observe that (6.5) relates $\{y_\alpha \mid \alpha \in F^r\}$ to $\{y_\alpha \mid \alpha \in E\}$, and remark that (6.5) always has a solution $\{y_\alpha \mid \alpha \in F^r\}$ no matter how $\{y_\alpha \mid \alpha \in E\}$ are specified (take $y_\alpha = \max_{\beta\in E} y_\beta$, for all $\alpha \in F^r$).

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

(a) *If* $y \in Y(v)$, *i.e. if* y *satisfies* (6.1), (6.2), (6.3) *it satisfies* (6.5) *as well.*

(b) *Conversely, if one picks* $\{y_\alpha \mid \alpha \in E\}$ *arbitrarily, next picks* $\{y_\alpha \mid \alpha \in F^r\}$ *to satisfy* (6.5), *next defines* $\{y_\alpha \mid \alpha \in F^t \cup F^\ell\}$ *as the right-hand side of* (6.4), *then the resulting vector* $\{y_\alpha \mid \alpha \in E \cup F\}$ *satisfies* (6.1), (6.2), (6.3), *hence belongs to* Y(v).

PROOF.

Part (a) follows from the above remarks.

(b) Observe that the right-hand side of (6.4) may be interpreted as the maximal total expected return of a terminating discrete-time Markovian model, with $F^t \cup F^\ell$ as state space. Because of the choice:

$$(6.6)\qquad y_\alpha = \max_{\phi\in\Phi} \sum_{\beta\in F^\ell\cup F^t} [I - W(\phi)]^{-1}_{\alpha\beta} [\tilde{q}(\phi)_\beta + \sum_{\gamma\in E\cup F^r} \tilde{P}(\phi)_{\beta\gamma} y_\gamma], \qquad \text{for } \alpha \in F^t \cup F^\ell,$$

it hence follows from corollary 2 of [21] that $y_\alpha = \tilde{q}(\phi)_\alpha + \sum_{\beta\in E\cup F^r} \tilde{P}(\phi)_{\alpha\beta} y_\beta + \sum_{\beta\in F^t\cup F^\ell} W(\phi)_{\alpha\beta} y_\beta$, $\alpha \in F^\ell$. Hence, the vector y satisfies (6.1) and (6.2)

In addition, using corollary 1 of [21], it follows that there exists a $\phi^* \in \Phi$ that maximizes the right-hand side of (6.6) simultaneously for all $\alpha \in F^t \cup F^\ell$, given any $\{y_\alpha \mid \alpha \in E \cup F^r\}$. Consider the inequalities (6.5) for $\phi = \phi^*$, and use (6.6) in order to show that the vector y satisfies (6.3) as well. □

REMARK 4. This provides a triangular decomposition in that one first determines $\{y_\alpha \mid \alpha \in E\}$, next $\{y_\alpha \mid \alpha \in F^r\}$ and finally $\{y_\alpha \mid \alpha \in F^\ell \cup F^t\}$. The last part can actually be decomposed further, by first determining $\{y_\alpha \mid \alpha \in F^\ell\}$ and then determining $\{y_\alpha \mid \alpha \in F^t\}$ via

$$y_\alpha = \max_{\phi\in\Phi} \sum_{\beta\in F^\ell} [I - W(\phi)^\ell]^{-1}_{\alpha\beta} [\tilde{q}(\phi)_\beta + \sum_{\gamma\in E} \tilde{P}(\phi)_{\beta\gamma} y_\gamma], \qquad \alpha \in F^\ell$$

$$y_\alpha = \max_{\phi\in\Phi} \sum_{\beta\in F^t} [I - W(\phi)^t]^{-1}_{\alpha\beta} [\tilde{q}(\phi)_\beta + \sum_{\gamma\in E\cup F^\ell\cup F^r} \tilde{P}(\phi)_{\beta\gamma} y_\gamma], \quad \alpha \in F^t,$$

where the transient matrices $W(\phi)^\ell$ and $W(\phi)^t$ are defined by:

$$W(\phi)^\ell \equiv [\tilde{P}(\phi)_{\alpha\beta}]_{\alpha,\beta\in F^\ell}; \qquad W(\phi)^t = [\tilde{P}(\phi)_{\alpha\beta}]_{\alpha,\beta\in F^t}.$$

Example 2 below has $N = 7$, $g_i^* = 0$, $L(i) = K(i)$ for all i
$R^* = \cup_{i=1}^{7} R^{*i}$ with $R^{*i} = \{i\}$, i.e. $n^* = 7$
$E = \{\alpha = 1\}$; $F^{\ell} = \{\alpha = 4\}$; $F^{t} = \{\alpha = 7\}$; $F^{r} = \{\alpha = 2,3,5,6\}$.
V is the solution set to the following decomposed set of inequalities:

$\alpha = 1$: v_1 arbitrary

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

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

$\alpha = (2,3)$: $q_2^2 \leq v_2 - v_3 \leq -q_3^2$

$\alpha = (5,6)$: $v_5 \geq q_5^2 + v_6,\ q_5^2 + q_6^3 + .5(v_1+v_2),\ q_5^2 + q_6^3 + .5q_2^2 + .r(v_1+v_3)$

$v_6 \geq q_6^2 + v_5,\ q_6^3 + .5(v_1+v_2),\ q_6^3 + .5q_2^2 + .5(v_1+v_2)$.

<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	0		0	1				
3	1	0		0	1				
	2	0		1	0				
4	1	0				1			
	2	0	1						
5	1	0					1	0	
	2	0					0	1	
6	1	0					0	1	
	2	0					1	0	
	3	0	.5	.5					
7	1	0							1
	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] BELLMAN, R., *A Markovian Decision Process*, J. Math. Mech. 6 (1957), 679-684.

[2] 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.

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

[4] 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.

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

[6] DENARDO, E., *Markov Renewal Programs with small interest rates*, Ann. of Math. Statistics 42 (1971), 477-496.

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

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

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

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

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

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

[13] LANERY, E., *Etude Asymptotique des Systemes Markoviens a Commande*, R.I.R.O. 1 (1967), 3-56.

[14] 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).

[15] 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.

[16] SCHWEITZER, P., *Perturbation theory and Markovian Decision Processes*, Ph.D. dissertation, MIT (1965) (MITORC report H15).

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

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

[19] SCHWEITZER, P. & A. FEDERGRUEN, *Asymptotic Value Iteration for Undiscounted Markov Decision Problems* (to appear).

[20] SCHWEITZER, P. & A. FEDERGRUEN, *Relative Values in the Policy Iteration Algorithm for Multichain Markov Renewal Programs* (to appear).

[21] VEINOTT, A. Jr., *Discrete Dynamic Programming with sensitive discount optimality criteria*, Ann. Math. Stat. 40 (1969), 1635-1660.

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