MATHEMATICS OF OPERATIONS RESEARCH
Vol. 2, No. 4, November 1977
Printed in U.S.A.

THE ASYMPTOTIC BEHAVIOR OF UNDISCOUNTED VALUE ITERATION IN MARKOV DECISION PROBLEMS*†

P. J. SCHWEITZER** AND A. FEDERGRUEN***

This paper considers undiscounted Markov Decision Problems. For the general multichain case, we obtain necessary and sufficient conditions which guarantee that the maximal total expected reward for a planning horizon of n epochs minus n times the long run average expected reward has a finite limit as $n \to \infty$ for each initial state and each final reward vector. In addition, we obtain a characterization of the chain and periodicity structure of the set of one-step and J-step maximal gain policies. Finally, we discuss the asymptotic properties of the undiscounted value-iteration method.

1. Introduction. The value-iteration equations for undiscounted Markov Decision Processes (MDPs) with finite state- and action space, were first studied by Bellman [2] and Howard [6]:

$$v(n+1)_i = Qv(n)_i, \qquad i = 1, \ldots, N, \tag{1.1}$$

where the Q operator is defined by:

$$Qx_i = \max_{k \in K(i)} \left\{ q_i^k + \sum_{j=1}^{N} P_{ij}^k x_j \right\}, \qquad i = 1, \ldots, N, \tag{1.2}$$

and $v(0)$ is a given N-vector. $\Omega = \{1, \ldots, N\}$ denotes the state space, $K(i)$ the finite set of alternatives in state i, q_i^k the one-step expected reward and $P_{ij}^k \geq 0$ the transition probability to state j, when alternative $k \in K(i)$ is chosen in state i ($i = 1, \ldots, N$).

For all $n = 1, 2, \ldots$ and $i \in \Omega$, $v(n)_i$ may be interpreted as the maximal total expected reward for a planning horizon of n epochs, when starting at state i and given an amount $v(0)_j$ is obtained when ending up at stage j. Bellman [2] showed that if every P_{ij}^k is strictly positive, then $v(n)_i \sim ng^*$, $n \to \infty$, the scalar g^* being the maximal gain rate and Howard [6] conjectured that there generally exist two N-vectors g^* and v^*, such that

$$\lim_{n \to \infty} v(n) - ng^* - v^* = 0. \tag{1.3}$$

Although Brown [3, theorem 4.3] showed that $v(n) - ng^*$ is bounded, provided g^* is taken as the maximal gain rate vector, the limit in (1.3) may not exist for arbitrary $v(0)$ if some of the transition probability matrices (*tpm's*) are periodic. The identification of sufficient conditions for the existence of the limit in (1.3) is of particular importance:

(a) when considering the infinite horizon-model with the average return per unit time criterion, as an approximation to the model where the planning horizon is finite though large.

* Received August 4, 1976; revised July 5, 1977.
AMS 1970 subject classification. Primary 90C40.
IAOR 1973 subject classification. Main: Markov Decision Programming. Cross Reference: Dynamic Programming.
Key words. Markov Decision Problems; average cost criterion, chain and periodicity structure, asymptotic behavior: value-iteration method.
† This paper is registered as Math. Center report BW 44/76.
** I.B.M. Thomas J. Watson Research Center and University of Rochester.
*** Foundation Mathematisch Centrum, Amsterdam.

(b) for the case $N \gg 1$, where the value-iteration method is the only practical way of locating maximal-gain policies. If the limit in (1.3) exists, then a generalization of Odoni [10] shows that any policy achieving the maxima in (1.1) for large n is maximal gain. However, if the limit in (1.3) fails to exist, then example 4 in Lanery [7] shows that policies achieving the maxima for large n in (1.1) need not be maximal gain.

Sufficiency conditions for the existence of the limit in (1.3) have been established by White [17] and Schweitzer [12], [13] in the unichain case, where $g_i^* = g^*$ (say) for all $i \in \Omega$.

Related convergence results for MDPs with compact action spaces, the denumerable and general state space case and for continuous time Markov Decision Processes were obtained in respectively Bather [1], Hordijk, Schweitzer and Tijms [5], Tijms [16] and Lembersky [8].

In this paper we establish the weakest sufficient condition. It holds for the general *multichain* case, and states that the limit in (1.3) exists for every $v(0) \in E^N$, *if and only if* there exists a randomized maximal gain policy whose tpm is aperiodic (but not necessarily unichained) and has $R^* = \{i \in \Omega \mid i \text{ is recurrent for some pure maximal gain policy}\}$ as its set of recurrent states.

In addition, we show that in general the sequence $\{v(n) - ng^*\}_{n=1}^{\infty}$ is asymptotically periodic, i.e. there exists an integer d^* (which merely depends upon the chain- and periodicity structure of the maximal gain policies), such that

$$\lim_{n\to\infty} v(nJ + r) - (nJ + r)g^* \text{ exists for all } v(0) \in E^N \tag{1.4}$$

if and only if J is a multiple of d^*.

The sufficiency parts of the above mentioned results were treated in Lanery [7]. However, it appears that the proof of proposition 19 in [7] from which the main result is derived, is either incomplete or incorrect (Note 1).

Moreover, our methods use the set of all randomized policies, and involve the analysis of the chain- and periodicity structure of the one- and J-step (randomized) maximal gain policies ($J \geq 1$). This enables a full characterization of the asymptotic period.

In §2, we give some notation and preliminaries. In §3, we analyze the periodicity structure of the maximal gain policies, while in §4 the chain- and periodicity-structure of the multi-step maximal gain policies is characterized. In §5, we obtain inter alia the above mentioned results with respect to the asymptotic periodicity, and the necessary and sufficient condition for the existence of the limit in (1.3) for all $v(0) \in E^N$.

Finally, we show how the behaviour of the various sequences $\{v(nJ + r)_i - (nJ + r)\cdot g_i^*\}_{n=1}^{\infty}$ ($r = 1, \ldots, J$; $i = 1, \ldots, N$) interdepends.

2. Notation and preliminaries. A (stationary) randomized policy f is a tableau $[f_{ik}]$ satisfying $f_{ik} \geq 0$ and $\sum_{k\in K(i)} f_{ik} = 1$, where f_{ik} is the probability that the kth alternative is chosen when entering state i.

We let S_R denote the set of all randomized policies, and S_P the set of all pure (nonrandomized) policies (i.e. each $f_{ik} = 0$ or 1). Associated with each $f \in S_R$, are an N-component reward vector $q(f)$ and $N \times N$-matrix $P(f)$:

$$q(f)_i = \sum_{k\in K(i)} f_{ik} q_i^k; \qquad P(f)_{ij} = \sum_{k\in K(i)} f_{ik} P_{ij}^k,\ 1 \leq i, j \leq N. \tag{2.1}$$

Note that $P(f)$ is a stochastic matrix ($P(f)_{ij} \geq 0$, $\sum_{j=1}^N P(f)_{ij} = 1$; $1 \leq i, j \leq N$). For any $f \in S_R$, we define the stochastic matrix $\Pi(f)$ as the Cesaro limit of the sequence $\{P^n(f)\}_{n=1}^{\infty}$, which always exists and has the following properties:

$$P(f)\Pi(f) = \Pi(f) = \Pi(f)P(f). \tag{2.2}$$

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

$$\Pi(f)_{ij} = \sum_{m=1}^{n(f)} \Phi_i^m(f)\pi^m(f)_j, \tag{2.3}$$

where $\pi^m(f)$ is the unique equilibrium distribution of $P(f)$ on the mth subchain $C^m(f)$, and $\Phi_i^m(f)$ is the probability of absorption in $C^m(f)$, starting from state i. Let $R(f) = \{j \mid \Pi(f)_{jj} > 0\}$, i.e. $R(f)$ is the set of recurrent states for $P(f)$.

Let $d^m(f) \geqslant 1$ denote the period of $C^m(f)$, and let $\{C^{m,\beta}(f) \mid \beta = 1, \ldots, d^m(f)\}$ indicate the set of cyclically moving subsets (*c.m.s.*) of $C^m(f)$ numbered such that for any $m = 1, \ldots, n(f)$ and $\beta = 1, \ldots, d^m(f)$ (cf. [11]):

$$i \in C^{m,\beta}(f) \Rightarrow P(f)_{ij} > 0 \text{ only if } j \in C^{m,\beta+1}(f) \tag{2.4}$$

with the convention that hereafter β in $C^{m,\beta}(f)$ is taken modulo $d^m(f)$ e.g. $C^{m,\beta+1}(f) = C^{m,1}(f)$ if $\beta = d^m(f)$.

For all $i \in C^m(f)$:
$d^m(f)$ = greatest common divisor (g.c.d.) of $\{n \mid P(f)^n_{ii} > 0\}$
= g.c.d. $\{n \mid \text{there exists a cycle } (s_0 = i, s_1, \ldots, s_n = i) \text{ for } P(f)\}$ (2.5)

where $(s_0 = i, s_1, \ldots, s_n = i)$ is called a *cycle* for $P(f)$ if $P(f)_{s_l s_{l+1}} > 0$ and if all the s_l are distinct ($l = 0, \ldots, n-1$).

$$\lim_{n\to\infty} P^{nd^m(f)+r}(f)_{ij} > 0, \quad \text{for all } i \in C^{m,\beta}(f) \text{ and } j \in C^{m,\beta+r}(f) \quad (r = 1, 2, \ldots). \tag{2.6}$$

For each $f \in S_R$, we define the gain rate vector $g(f) = \Pi(f)q(f)$, such that $g(f)_i$ represents the long run average expected return per unit time, when the initial state is i, and policy f is used. We thus have

$$g(f)_i = \sum_{m=1}^{n(f)} \Phi_i^m(f)g^m(f), \qquad i \in \Omega, \tag{2.7}$$

with

$$g^m(f) = \langle \pi^m(f), q(f)\rangle, \qquad m = 1, \ldots, n(f).$$

Next define:

$$g_i^* = \sup_{f\in S_R} g(f)_i; \qquad i = 1, \ldots, N. \tag{2.8}$$

Since Derman [4] proved the existence of pure policies f which attain the N suprema in (2.8) simultaneously, we can define:

$$S_{PMG} = \{f \in S_P \mid g(f) = g^*\}; \qquad S_{RMG} = \{f \in S_R \mid g(f) = g^*\} \tag{2.9}$$

as the set of all pure, and the set of all randomized, maximal gain policies.

Finally define R^* as the set of states that are recurrent under some maximal gain policy:

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

The following lemma which was proved in Schweitzer and Federgruen [14, theorem 3.2] provides a basic characterization of this set:

LEMMA 2.1. (a) $R^* = \{i \mid i \in R(f) \text{ for some } f \in S_{PMG}\}$.
(b) *The set* $\{f \in S_{RMG} \mid R(f) = R^*\}$ *is not empty.*
(c) *Define* $n^* = \min\{n(f) \mid f \in S_{RMG} \text{ with } R(f) = R^*\}$ *and* $S^*_{RMG} = \{f \in S_{RMG} \mid R(f) = R^* \text{ and } n(f) = n^*\}$.
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* $\{R^{*\alpha}, \alpha = 1, \ldots, n^*\}$.
(e) *For any* $\alpha \in \{1, \ldots, n^*\}$, $g_i^* = g^{*\alpha}$ *(say) for all* $i \in R^{*\alpha}$.
(f) *Let* $R^{(1)}, \ldots, R^{(m)}$ *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, $1 \leqslant k \leqslant m$;
(2) *there exists an* $f \in S_{RMG}$, *with* $\{R^{(k)} \mid k = 1, \ldots, m\}$ *as its set of subchains.*
Then $m = n^*$ *and after renumbering* $R^{(\alpha)} = R^{*\alpha}$, $\alpha = 1, \ldots, n^*$.

Define the operator T by

$$Tx_i = \max_{k \in L(i)} \left\{ q_i^k + \sum_j P_{ij}^k x_j \right\}, \quad i = 1, \ldots, N, \tag{2.10}$$

where

$$L(i) = \left\{ k \in K(i) \mid g_i^* = \sum_j P_{ij}^k g_j^* \right\}, \quad \text{for all } i \in \Omega.$$

Let Q^n(and T^n) denote the n-fold application of the operator $Q(T)$:

$$Q^n x = Q(Q^{n-1}x); \quad T^n x = T(T^{n-1}x); \qquad n = 2, 3, \ldots \text{ and } x \in E^N$$
$$(\text{with } Q^1 x = Qx \text{ and } T^1 x = Tx).$$

The basic properties of both operators were studied in Schweitzer and Federgruen [15]. In particular, it was shown that the Q operator reduces to T in the following two ways:

for each $x \in E^N$, there exists a scalar $t_0(x)$, such that $Q^n(x + tg^*) = T^n(x + tg^*)$ for $n = 1, 2, \ldots$ and $t \geqslant t_0(x)$ (cf. [15, lemma 2.2 part (c)]), (2.11)

for each $x \in E^N$ there exists an integer $n_0(x)$ such that $Q^{n+1}x = T(Q^n x) = T^{n+1-n_0(x)} Q^{n_0(x)} x$, for all $n \geqslant n_0(x)$ (cf. [3] and [15, lemma 2.2 part (c)]). (2.12)

We next consider the functional equation:

$$v + g^* = Tv. \tag{2.13}$$

Let $V = \{v \in E^N \mid v \text{ satisfies (2.13)}\}$ and define for any $v \in V$:

$$b(v)_i^k = q_i^k - g_i^* + \sum_{j=1}^N P_{ij}^k v_j - v_i, \qquad i \in \Omega,\ k \in K(i),$$

$$b(v, f)_i = \sum_{k \in K(i)} f_{ik} b(v)_i^k = \left[q(f) - g^* + P(f)v - v \right]_i, \qquad i \in \Omega,\ f \in S_R. \tag{2.14}$$

Observe that for all $v \in V$, $\max_{k \in L(i)} b(v)_i^k = 0$, for all $i \in \Omega$. Finally, we define for any $i \in R^*$, the set $K^*(i)$ as the set of actions which a pure maximal gain policy that has i among its recurrent states, could prescribe:

$$K^*(i) = \{k \in K(i) \mid \text{there exists an } f \in S_{PMG}, \text{ with } i \in R(f) \text{ and } f_{ik} = 1\}. \tag{2.15}$$

The following lemma gives the necessary and sufficient condition for a policy to be maximal gain, characterizes the sets $K^*(i)$ and shows that any policy that randomizes among *all* actions in $K^*(i)$, in each of the states in R^*, and among *all* actions in $L(i)$ for the states in $\Omega - R^*$, belongs to S^*_{RMG}:

LEMMA 2.2. (a) *Fix $v \in V$. A policy $f \in S_R$ is maximal gain (i.e. $f \in S_{RMG}$) if and only if*

(1) *for all $i \in \Omega$, $f_{ik} > 0 \Rightarrow k \in L(i)$, i.e. $P(f)g^* = g^*$;*

(2) *for all $i \in R(f)$, $f_{ik} > 0 \Rightarrow b(v)_i^k = 0$, i.e. $\Pi(f)b(v, f) = 0$.*

(b) *$K^*(i) = \{k \in L(i) \mid$ there exists an $f \in S_{RMG}$, with $i \in R(f)$, and $f_{ik} > 0\}$, $i \in R^*$.*

(c) *For any $v \in V$, $K^*(i) = \{k \in L(i) \mid b(v)_i^k = 0$ and $\sum_{j \in R^{*\alpha}} P_{ij}^k = 1\}$, for all $i \in R^{*\alpha}$, $\alpha = 1, \ldots, n^*$.*

(d) *Define $f^* \in S_R$ such that*

$$\{k \mid f^*_{ik} > 0\} = \begin{cases} K^*(i), & i \in R^*, \\ L(i), & i \in \Omega - R^*. \end{cases}$$

Then $f^* \in S^*_{RMG}$.

PROOF. (a) cf. theorem 3.1, part (a) in [14].

(b) Clearly, $K^*(i)$ is contained within the set on the right-hand side. Next, fix $i \in R^*$, $k \in K(i)$ and $f \in S_{RMG}$, such that $i \in R(f)$ and $f_{ik} > 0$, and use lemma 2.1 in [13] in order to show that there exists an $h \in S_{PMG}$, with $i \in R(h)$, and $h_{ik} = 1$ as well, which proves the reversed inclusion.

(c) Fix $\alpha \in \{1, \ldots, n^*\}$, $i_0 \in R^{*\alpha}$. First, let $k \in K^*(i)$ and $f \in S_{RMG}$, with $i \in R(f)$ and $f_{ik} > 0$, and apply part (a) of this lemma, and part (c) of lemma 2.1, in order to prove that $K^*(i)$ is contained within the set on the right-hand side of the equality. Next, take $k_0 \in L(i_0)$ such that $b(v)_{i_0}^{k_0} = 0$ and $\sum_{j \in R^{*\alpha}} P_{i_0 j}^{k_0} = 1$, and fix $f^* \in S^*_{RMG}$. Define f^{**} such that

$$f^{**}_{i_0 k_0} = 1, \quad \text{and} \quad f^{**}_{jk} = f^*_{jk}, \quad \text{for all } j \neq i_0,\ k \in K(j).$$

Use part (d) of lemma 2.1, in order to show that all states in $R^{*\alpha} \setminus \{i_0\}$ can reach state i_0 under $P(f^{**})$ whereas state i_0 can only reach states within $R^{*\alpha}$. We conclude that $i_0 \in R(f^{**})$, while $f^{**} \in S_{RMG}$, as can be verified using part (a) of this lemma, thus proving the reversed inclusion.

(d) Cf. remark 1 in [14]. ∎

We finally need the following lemma:

LEMMA 2.3. (a) *Fix $f^1, f^2 \in S_R$, and let C^1 and C^2 be two subchains of $P(f^1)$ and $P(f^2)$ with period d^1 and d^2 respectively, such that $C^1 \cap C^2 \neq \emptyset$. Define f^3 such that*

$$\{k \mid f^3_{ik} > 0\} = \begin{cases} \{k \mid f^2_{ik} > 0\} & \text{for all } i \in C^2 \setminus C^1, \\ \{k \mid f^1_{ik} > 0\} \cup \{k \mid f^2_{ik} > 0\} & \text{for all } i \in C^1 \cap C^2, \\ \{k \mid f^1_{ik} > 0\} & \text{otherwise.} \end{cases}$$

Then

(1) *$C^1 \cup C^2$ is a subchain of $P(f^3)$, the period d^3 of which is a common divisor of d^1 and d^2.*

(2) *If $f^1, f^2 \in S_{RMG}$, then $f^3 \in S_{RMG}$.*

(b) *For any $f \in S_R$, define the set of pure policies $S_P(f) = X_{i \in \Omega}\{k \mid f_{ik} > 0\}$. Then for all $m = 1, \ldots, n(f)$:*

$$d^m(f) = \text{g.c.d.}\{d^r(h) \mid h \in S_P(f),\ 1 \leqslant r \leqslant n(h),\ C^r(h) \subseteq C^m(f)\}. \quad (2.16)$$

PROOF. (a) (1) Show that $C^1 \cup C^2$ is a closed and communicating set of states for $R(f^3)$. The former is immediate; the latter holds since any state in $C^1 \cap C^2$ communicates with $C^1 \cup C^2$. Fix $i \in C^1 \cap C^2$. Since $\{n \mid P(f^3)^n_{ii} > 0\} \supseteq \{n \mid P(f^1)^n_{ii} > 0\} \cup \{n \mid P(f^2)^n_{ii} > 0\}$, it follows (cf. (2.5)) that $d^3 = \text{g.c.d.}\{n \mid P(f^3)^n_{ii} > 0\}$ is a common divisor of d^1 and d^2.

(2) Observe that for each $i \in \Omega$, $f^3_{ik} > 0$ only for $k \in L(i)$ since it follows from lemma 2.2 part (a) that $f^1_{ik} > 0$ and $f^2_{ik} > 0$ only for $k \in L(i)$. Using the fact that $R(f^3) \subseteq R(f^1) \cup C^2$, and applying lemma 2.2 part (a2) one verifies that $f^3 \in S_{RMG}$.

(b) Fix $m \in \{1, \ldots, n(f)\}$ and $h \in S_P(f)$. Since $C^m(f)$ is closed under any policy in $S_P(f)$, $P(h)$ has a subchain $C^r(h) \subseteq C^m(f)$ $(1 \leqslant r \leqslant n(h))$. Since $P(h)_{ij} > 0$ only if $P(f)_{ij} > 0$, and since $i \in C^m(f)$ implies that $P(f)^t_{ii} > 0$ only if t is a multiple of $d^m(f)$, it follows that for $i \in C^r(h)$, $P(h)^t_{ii} > 0$ only if t is a multiple of $d^m(f)$. Thus (2.5) implies that the left-hand side of (2.16) is less than or equal to its right-hand side. To prove the reversed inequality in (2.16) fix $i \in C^m(f)$ and recall from (2.5) that

$$d^m(f) = \text{g.c.d.}\{n \mid \text{there exists a cycle } (s_0 = i, \ldots, s_n = i) \text{ of } P(f)\}. \tag{2.17}$$

We next show that

$$\text{for each cycle } S = \{s_0 = i, s_1, \ldots, s_n = i\} \text{ of } P(f), \text{ there exists a pure policy } h \in S_P(f) \text{ which has } i \text{ recurrent and contains the same cycle.} \tag{2.18}$$

As a consequence, we obtain that each of the elements in the set to the right of (2.17) is a multiple of the period of a subchain of a pure policy that lies within $C^m(f)$, thus proving the reversed inequality in (2.16) and hence part (b).

In order to show (2.18), construct the policy $h \in S_P(f)$ as follows: Let $h_{s_l k} = 1$ for any one k such that $f_{s_l k} > 0$ and $P^k_{s_l s_{l+1}} > 0$ $(l = 0, \ldots, n-1)$; for $j \notin C^m(f)$, let $h_{jk} = 1$ for any one k such that $f_{jk} > 0$. If $S \neq C^m(f)$, let Δ initially be equal to S, and define $\bar{\Delta} = C^m(f) \backslash \Delta$.

Next, the following step is performed: Choose a state $j \in \bar{\Delta}$ and an alternative k such that $f_{jk} > 0$ and $P^k_{jt} > 0$ for some $t \in \Delta$, transfer j from $\bar{\Delta}$ to Δ and define $h_{jk} = 1$. Such k and t can always 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 S is a cycle for $P(h)$, with $i \in R(h)$ since i can be reached from any state in $C^m(f)$, and $C^m(f)$ is closed under $P(h)$. ∎

REMARK 1. The period d^3, defined in part (a) of the previous lemma, does not necessarily have to be the greatest common divisor of d^1 and d^2. Take

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

with $d^1 = d^2 = 3$ and $d^3 = 1$. However, it can be shown that $d^3 = \text{g.c.d.}\{d^1, d^2\}$ does hold when $P(f^1)$ and $P(f^2)$ merely differ in one row, the corresponding state being recurrent for both chains (cf. part (b)).

3. The periodicity structure of the policies in S_{RMG}.

We first define

$$d(\alpha) = \min\{d^m(f) \mid f \in S_{RMG}, 1 \leqslant m \leqslant n(f), C^m(f) \subseteq R^{*\alpha}\}, \qquad \alpha = 1, \ldots, n^*, \tag{3.1}$$

$$d_i = \min\{d^m(f) \mid f \in S_{RMG}, 1 \leqslant m \leqslant n(f), i \in C^m(f)\}, \qquad i \in R^*, \tag{3.2}$$

i.e. $d(\alpha)$ $[d_i]$ denotes the minimum of the periods of the subchains of the maximal gain policies that lie within $R^{*\alpha}$ [that contain the state i]. Let $f^* \in S^*_{RMG}$ be defined as in lemma 2.2 part (d), i.e. let

$$\{k \mid f^*_{ik} > 0\} = \begin{cases} K^*(i), & i \in R^*, \\ L(i), & i \in \Omega \backslash R^*. \end{cases}$$

For each $\alpha = 1, \ldots, n^*$ and $t = 1, \ldots, d^\alpha(f^*)$ let $R^{*\alpha, t} = C^{\alpha, t}(f^*)$ with the convention that hereafter t in $R^{*\alpha, t}$ is taken modulo $d^\alpha(f^*)$ (e.g. $R^{*\alpha, t} = R^{*\alpha, 1}$ if $t = d^\alpha(f^*) + 1$).

THEOREM 3.1 (PERIODICITY STRUCTURE) (CF. LEMMA 2.1). (a) $d^\alpha(f^*) = d(\alpha)$, $\alpha = 1, \ldots, n^*$.

(b) *Fix* $\alpha \in \{1, \ldots, n^*\}$. *Let* $h \in S_{RMG}$ *and* $C^m(h) \subseteq R^{*\alpha}$. *Then* $d^m(h)$ *is a multiple of* $d(\alpha)$.

(c) $d(\alpha) = \text{g.c.d.}\{d^m(f) \mid f \in S_{PMG},\ 1 \leqslant m \leqslant n(f),\ C^m(f) \subseteq R^{*\alpha}\}$, $\alpha = 1, \ldots, n^*$.

(d) $d_i = d(\alpha)$ *for all* $i \in R^{*\alpha}$, $\alpha = 1, \ldots, n^*$.

(e) $d(\alpha) = \min\{d^\alpha(f) \mid f \in S^*_{RMG}\}$, $\alpha = 1, \ldots, n^*$.

(f) *The set* $S^{**}_{RMG} = \{f \in S^*_{RMG} \mid d^\alpha(f) = d(\alpha),\ \alpha = 1, \ldots, n^*\}$ *is nonempty.*

(g) *For each* $i \in R^*$, *say* $i \in R^{*\alpha, t}$ $(1 \leqslant \alpha \leqslant n^*;\ 1 \leqslant t \leqslant d(\alpha))$ *and* $k \in K^*(i)$: $P^k_{ij} > 0 \Rightarrow j \in R^{*\alpha, t+1}$.

(h) *For each* $h \in S_{RMG}$, *and* $i \in R(h) \cap R^{*\alpha, t}$ $(1 \leqslant \alpha \leqslant n^*;\ 1 \leqslant t \leqslant d(\alpha))$ $P(h)_{ij} > 0$ *only for* $j \in R^{*\alpha, t+1} \cap R(h)$.

(i) *Fix* $h \in S_{RMG}$, *with* $C^m(h) \subseteq R^{*\alpha}$ $(1 \leqslant m \leqslant n(h);\ 1 \leqslant \alpha \leqslant n^*)$. $C^m(h)$ *has* $d^m(h) / d(\alpha)$ *c.m.s. within each of the sets* $R^{*\alpha, t}$ $(1 \leqslant t \leqslant d(\alpha))$.

(j) *All* $f \in S^{**}_{RMG}$ *have the same collection of c.m.s.* $\{R^{*\alpha, t} \mid \alpha = 1, \ldots, n^*;\ t = 1, \ldots, d(\alpha)\}$.

(k) *Let* $R^{(1)}, \ldots, R^{(M)}$ *be disjoint sets of states, such that*

(1) *If* C *is a c.m.s. of some subchain of some* $f \in S_{RMG}$, *then* $C \subseteq R^{(k)}$ *for some* k, $1 \leqslant k \leqslant M$.

(2) *There exists a* $f \in S_{RMG}$, *with* $\{R^{(k)} \mid k = 1, \ldots, M\}$ *as its collection of c.m.s. Then* $M = \sum_{\alpha=1}^{n^*} d(\alpha)$ *and there is a one-to-one correspondence between the sets* $\{R^{(k)} \mid K = 1, \ldots, M\}$ *and the sets* $\{R^{*\alpha, t} \mid \alpha = 1, \ldots, n^*;\ t = 1, \ldots, d(\alpha)\}$.

PROOF. (a),(b) Fix $\alpha \in \{1, \ldots, n^*\}$ and let $h \in S_{RMG}$, with $C^m(h) \subseteq R^{*\alpha}$ (for some m, $1 \leqslant m \leqslant n(h)$). Define f^{**} such that

$$\{k \mid f^{**}_{ik} > 0\} = \begin{cases} \{k \mid h_{ik} > 0\} \cup \{k \mid f^*_{ik} > 0\}, & \text{for all } i \in C^m(h), \\ \{k \mid f^*_{ik} > 0\}, & \text{otherwise.} \end{cases}$$

It then follows from the definitions of the policy f^* and the sets $K^*(i)$ (cf. lemma 2.2 part (b)) that

$$\{k \mid f^{**}_{ik} > 0\} = \begin{cases} K^*(i) & \text{for } i \in R^*, \\ L(i) & \text{for } i \in \Omega \backslash R^* \end{cases}$$

which implies that f^* and f^{**} have the same chain- and periodicity structure. In particular, $d^\alpha(f^{**}) = d^\alpha(f^*)$. On the other hand, applying lemma (2.3), part (a), it follows that $d^\alpha(f^{**})$ is a divisor of $d^m(h)$, hence

$$d^\alpha(f^*) \text{ divides } d^m(h), \tag{3.3}$$

so that

$$d(\alpha) \leqslant d^\alpha(f^*) \leqslant \min\{d^m(h) \mid h \in S_{RMG},\ 1 \leqslant m \leqslant n(h),\ C^m(h) \subseteq R^{*\alpha}\} = d(\alpha).$$

This proves part (a), whereas the combination of part (a) and (3.3) proves part (b).

(c) Use the fact that $d(\alpha) = d^{\alpha}(f^*)$; apply lemma (2.3) part (b), and use the definition of the sets $K^*(i)$.

(d) Fix $i \in R^{*\alpha}$. Clearly $d_i \geq d(\alpha)$ (cf. (3.1) and (3.2)) and use part (a) to show $d_i \leq d(\alpha)$ as well.

(e),(f) immediate from part (a).

(g) Observe that $P(f^*)_{ij} > 0 \Rightarrow j \in R^{*\alpha, t+1}$ (cf. (2.4)) and use lemma 2.2 part (d).

(h) Use the fact that $h_{ik} > 0$ only for $k \in K^*(i)$ (cf. lemma 2.2 part (b)) and apply part (g).

(i) Recall from part (b) that $d^m(h)$ is a multiple of $d(\alpha)$. Take $i \in C^{m,1}(h)$, assume $i \in R^{*\alpha, t}$ $(1 \leq t \leq d(\alpha))$ and fix $s \in \{0, \ldots, d(\alpha) - 1\}$. In view of part (h), we obtain for $r = 0, \ldots, d^m(h)/d(\alpha) - 1$:

$$P(h)_{ij}^{nd^m(h)+rd(\alpha)+s} > 0 \quad \text{only for } j \in R^{*\alpha, t+s};\ n = 1, 2, \ldots .$$

Since $\lim_{n\to\infty} P(h)_{ij}^{nd^m(h)+rd(\alpha)+s} > 0$ for all $j \in C^{m, rd(\alpha)+s+1}(h)$ (cf.(2.4)) we conclude that $C^{m, rd(\alpha)+s+1}(h) \subseteq R^{*\alpha, t+s}$ for $r = 0, \ldots, d^m(h)/d(\alpha) - 1$ which proves part (i).

(j) Let $f \in S^{**}_{RMG}$ and fix $\alpha \in \{1, \ldots, n^*\}$. It follows from part (i) that each of the sets $R^{*\alpha, t}(1 \leq t \leq d(\alpha))$ contains exactly *one* c.m.s. $C^{\alpha, s}(f)$ (for some $1 \leq s \leq d(\alpha)$) of $P(f)$.

Since $R^{*\alpha} = \bigcup_{s=1}^{d(\alpha)} C^{\alpha, s}(f) = \bigcup_{t=1}^{d(\alpha)} R^{*\alpha, t}$, we conclude that for any $1 \leq s \leq d(\alpha)$:

$$C^{\alpha, s}(f) = R^{*\alpha, t} \quad \text{for some } t = t(s)$$

which proves that all $f \in S^{**}_{RMG}$ have the same collection of c.m.s.

(k) Apply property (1) to conclude that $R^{*\alpha, t} \subseteq R^{(k(\alpha, t))}$ for $\alpha = 1, \ldots, n^*$; $t = 1, \ldots, d(\alpha)$, and apply property (2) and part (i) to conclude

$$R^{(k)} \subseteq \text{some } R^{*\alpha, t}, \qquad k = 1, \ldots, M. \quad \blacksquare$$

REMARK 2. In [14], a finite procedure was given for calculating R^*, n^* and each $R^{*\alpha}$ after using the Policy Iteration Algorithm to find g^* and a $v \in V$.

Part (a) of the previous theorem shows that this procedure can be extended in order to find the $d(\alpha)$, the sets $R^{*\alpha, \beta}$ and a $f \in S^{**}_{RMG}$ in a finite number of calculations, as well:

(1) For each $i \in R^*$, determine the sets $K^*(i)$ (use lemma 2.2 part (c)).

(2) Define $f^* \in S^{**}_{RMG}$ by

$$\{k \mid f^*_{ik} > 0\} = \begin{cases} K^*(i), & i \in R^*, \\ L(i), & i \in \Omega \backslash R^*. \end{cases}$$

Then the cyclically moving subsets of each subchain $R^{*\alpha}$ of $P(f^*)$ form the $\{R^{*\alpha, t}\}_{t=1}^{n^*}$.

Consider the following example:

EXAMPLE 1.

i	k	P_{i1}^k	P_{i2}^k	P_{i3}^k	P_{i4}^k	P_{i5}^k	q_i^k
1	1	1	0	0	0	0	0
2	1	0	0	1	0	0	$q_2^1 \leq 0$
	2	0	0	0	1	0	0
3	1	0	1	0	0	0	0
4	1	0	0	0	0	1	0
5	1	0	0	1	0	0	0
	2	0	1	0	0	0	$q_5^2 \leq 0$
	3	1	0	0	0	0	0

Table 1 lists the six pure policies, their subchains and periods. Observe that (whatever the specific value of q_2^1, q_5^2): $g^* = (0, 0, 0, 0, 0)$; $K(i) = L(i)$ for all $i \in \Omega$ and $V = \{(x_1, \ldots, x_5) \mid x_1 \leqslant x_2 = x_3 = x_4 = x_5\}$, $n^* = 2$, $R^{*1} = \{1\}$; $R^{*2} = \{2, 3, 4, 5\}$; since $d(1) = 1$, $R^{*1,1} = \{1\}$.

TABLE 1

f	$S_{P(f)}$	$n(f)$	$C^1(f)$	$C^2(f)$	$d^1(f)$	$d^2(f)$
f_1	(1, 1, 1, 1, 1)	2	{1}	{2, 3}	1	2
f^2	(1, 1, 1, 1, 2)	2	{1}	{2, 3}	1	2
f^3	(1, 1, 1, 1, 3)	2	{1}	{2, 3}	1	2
f^4	(1, 2, 1, 1, 1)	2	{1}	{2, 3, 4, 5}	1	4
f^5	(1, 2, 1, 1, 2)	2	{1}	{2, 4, 5}	1	3
f^6	(1, 2, 1, 1, 3)	1	{1}	—	1	—

Next, consider the following cases:

case	q_2^1	q_5^2	S_{PMG}	$K^*(2)$	$K^*(5)$
1	0	0	$\{f^1, f^2, f^3, f^4, f^5, f^6\}$	{1, 2}	{1, 2}
2	<0	0	$\{f^4, f^5, f^6\}$	{2}	{1, 2}
3	0	<0	$\{f^1, f^2, f^3, f^4, f^6\}$	{1, 2}	{1}
4	<0	<0	$\{f^4, f^6\}$	{2}	{1}

Define $f^* \in S_R$ as in lemma 2.2 part (d):

$$P(f^*) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & x & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & x & x & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & x & x & 0 & 0 \end{bmatrix}$$

case 1 case 2

$$P(f^*) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & x & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

case 3 case 4

In case 1, $P(f^*)$ is aperiodic and $d(2) = 1 = \text{g.c.d.}\{2, 2, 2, 4, 3\}$ (cf. theorem 3.1(a), (c)). In case 2, $P(f^*)$ is aperiodic and $d(2) = 1 = \text{g.c.d.}\ \{4, 3\}$ (cf. theorem 3.1(a), (c)). In case 3, $P(f^*)$ has R^{*2} periodic with $d(2) = 2 = \text{g.c.d.}\ \{2, 2, 2, 4\}$ (cf. theorem 3.1(a), (c)).

$$R^{*2,1} = \{2, 5\}; \qquad R^{*2,2} = \{3, 4\}.$$

In case 4, $P(f^*)$ has R^{*2} periodic with $d(2) = 4 = \text{g.c.d.}\ \{4\}$ (cf. theorem 3.1(a), (c)).

$$R^{*2,1} = \{2\}; \quad R^{*2,2} = \{4\}; \quad R^{*2,3} = \{5\}; \quad R^{*2,4} = \{3\}.$$

Thus randomization is essential for both the recurrency properties and the periodicity structure: it plays the indispensable role of coalescing subchains and of decreasing periods. 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, or which achieves the minimal period in every subchain. For instance, case 1 of example 1 with state 1 and actions 1 and 3 in state 5 omitted, shows that

(a) all pure (maximal gain) policies have periodic tpm's, while a randomized (maximal gain) policy is aperiodic.

(b) none of the pure (maximal gain) policies has R^* as its recurrent set, although a randomized (maximal gain) policy does.

Observe that while $d(\alpha) = \text{g.c.d.}\ \{d^m(f) \mid f \in S_{PMG},\ 1 \leqslant m \leqslant n(f),\ C^m(f) \subseteq R^{*\alpha}\}$ for all $\alpha = 1, \dots, n^*$ (cf. part (a) of theorem 3.1), we may have

$$d_i = \text{g.c.d.}\ \{d^m(f) \mid f \in S_{RMG},\ 1 \leqslant m \leqslant n(f),\ i \in C^m(f)\}$$
$$< \text{g.c.d.}\ \{d^m(f) \mid f \in S_{PMG},\ 1 \leqslant m \leqslant n(f),\ i \in C^m(f)\}.$$

(Take case 1 of example 1, and $i = 3$.)

4. The multi-step policies. Fix an integer $J \geqslant 2$, and observe from (1.2) that

$$Q^J x_i = \max_{\xi \in \tilde{K}(i)} \left\{ \tilde{q}_i^\xi + \sum_j \tilde{P}_{ij}^\xi x_j \right\}$$

where $\tilde{K}(i) = \{(f^1, \dots, f^J) \mid f^1, \dots, f^J \in S_P\}$,

$$\tilde{q}_i^\xi = q(f^1)_i + P(f^1)q(f^2)_i + \cdots + \left[P(f^1) \dots P(f^{J-1})\right]q(f^J)_i,$$
$$i \in \Omega,\ \xi = (f^1, \dots, f^J) \in \tilde{K}(i),$$
$$\tilde{P}_{ij}^\xi = P(f^1) \cdots P(f^J)_{ij};\ 1 \leqslant i, j \leqslant N \text{ and}$$
$$\xi = (f^1, \dots, f^J) \in \tilde{K}(i). \tag{4.1}$$

Let $\tilde{Q} = Q^J$, and define a related "J-step"-MDP, denoted by a tilde, with Ω as its state space, $\tilde{K}(i)$ as the (finite) set of alternatives in state $i \in \Omega$, $\tilde{q}_i^\xi$ as the one-step expected reward and $\tilde{P}_{ij}^\xi$ as the transition probability to state j, when alternative $\xi \in \tilde{K}(i)$ is chosen when entering state i.

Let $\tilde{S}_R$ denote the set of all (stationary) randomized policies with respect to the above defined MDP, and observe that

$$\tilde{S}_R = \underset{i \in \Omega}{\times} \underset{r=1}{\overset{J}{\times}} S_R.$$

In complete analogy to the definitions given in §2, we define the operator $\tilde{T}$, the sets $\tilde{S}_P$, $\tilde{S}_{PMG}$, $\tilde{S}_{RMG}$, $\tilde{S}_{RMG}^*$, $\tilde{S}_{RMG}^{**}$, $\tilde{R}^*$, $\tilde{R}^{*\alpha}$, $\tilde{R}^{*\alpha,\beta}$, $\tilde{V}$, the integers $\tilde{n}^*$, $\tilde{d}(\alpha)$, $\tilde{d}_i$, and for each $\phi \in \tilde{S}_R$, the quantities $\tilde{q}(\phi)$, $\tilde{P}(\phi)$, $\tilde{\Pi}(\phi)$, $\tilde{g}(\phi)$, $\tilde{n}(\phi)$, $\tilde{d}^m(\phi)$, and for each $i \in \Omega$, the set $\tilde{L}(i)$.

Observe that a "J-step policy" $\phi \in \tilde{S}_R$ is specified by NJ "one-step" policies $\{\phi^{r,i} \mid r = 1, \dots, J;\ i = 1, \dots, N\}$ such that policy ϕ uses "action" $(\phi^{1,i}, \phi^{2,i}, \dots, \phi^{J,i}) \in \tilde{K}(i)$ while in state $i \in \Omega$:

$$\tilde{q}(\phi)_i = q(\phi^{1,i})_i + P(\phi^{1,i})q(\phi^{2,1})_i + \cdots + \left[P(\phi^{1,i}) \dots P(\phi^{J-1,i})\right]q(\phi^{J,i})_i,$$
$$\tilde{P}(\phi)_{ij} = P(\phi^{1,i}) \cdots P(\phi^{J,i})_{ij}, \qquad i, j \in \Omega.$$

The following theorem characterizes the "J-step" maximal gain policies and shows how their chain- and periodicity structure are connected with the corresponding ones in our original MDP.

First, define for any $\phi \in \tilde{S}_R$:

$$T^{r,i}(\phi) = \{j \mid P(\phi^{1,i}) \dots P(\phi^{r,i})_{ij} > 0\}, \qquad i \in \Omega, r = 1, \dots, J,$$
$$T^{0,i}(\phi) = \{i\}, \qquad i \in \Omega. \tag{4.2}$$

THEOREM 4.1. *Fix $J \geqslant 2$. Then*

(a) $\tilde{g}^* = Jg^*$ *and* $\{\phi \mid$ *there exists* $f \in S_{RMG}$ *such that* $\phi^{r,i} = f$ *for all* $r = 1, \dots, J$; $i = 1, \dots, N\} \subseteq \tilde{S}_{RMG}$.

(b) *Fix* $i \in \Omega$. *Let* $\xi = (f^1, \dots, f^J) \in K(i)$. *The following statements are equivalent*:

(1) $\xi \in \tilde{L}(i)$.

(2) $f^1_{ik} = 1 \Rightarrow k \in L(i)$.

$f^r_{jk} = 1 \Rightarrow k \in L(j)$ *for* $2 \leqslant r \leqslant J$ *and all* j *such that* $P(f^1) \dots P(f^{r-1})_{ij} > 0$.

(c) V *is an* n^*-*dimensional subset of the* $\tilde{n}^*$-*dimensional set* $\tilde{V}$.

(d) *Fix* $v \in V$. *Then* $\phi \in \tilde{S}_{RMG}$ *if and only if*

$$\phi^{r+1,}_{jk} > 0 \Rightarrow k \in L(i), \quad \text{for all } j \in T^{r,i}(\phi), i \in \Omega, r = 0, \dots, J-1,$$
$$b(v, \phi^{r+1,i})_j = 0 \quad \text{for all } j \in T^{r,i}(\phi), i \in \tilde{R}(\phi), r = 0, \dots, J-1. \tag{4.3}$$

(e) *Fix* $f \in S^{**}_{RMG}$, *and take* $\phi \in \tilde{S}_R$ *such that* $\phi^{i,r} = f$ *for all* $i \in \Omega$, $r = 1, \dots, J$. *Then*

(1) $\tilde{R}(\phi) = R^*$.

(2) *The collection of subchains of* $\tilde{P}(\phi)$ *is given by*:

$$\left\{ \bigcup_{k=1}^{\infty} R^{*\alpha, r+kJ} \mid \alpha = 1, \dots, n^*; r = 1, \dots, \text{g.c.d.}(J, d(\alpha)) \right\}. \tag{4.4}$$

(3) *Each of the* $R^{*\alpha, t}$ ($\alpha = 1, \dots, n^*$; $t = 1, \dots, d(\alpha)$) *is a cyclically moving subset of* $\tilde{P}(\phi)$.

(f) $\tilde{R}^* = R^*$.

(g) $\{\tilde{R}^{*\gamma} \mid \gamma = 1, \dots, \tilde{n}^*\} = \{\bigcup_{k=1}^{\infty} R^{*\alpha, r+kJ} \mid \alpha = 1, \dots, n^*;\ r = 1, \dots, \text{g.c.d.}(J, d(\alpha))\}$ *i.e.* $\tilde{n}^* = \sum_{\alpha=1}^{n^*} \text{g.c.d.}(J, d(\alpha)) \geqslant n^*$.

(h) $\{\tilde{R}^{*\alpha, t}\} = \{R^{*\alpha, t}\}$; *i.e. fix* $\alpha \in \{1, \dots, n^*\}$; *then* $\tilde{d}(\beta) = d(\alpha)/\text{g.c.d.}(J, d(\alpha))$ *for all* $\tilde{R}^{*\beta} \subseteq R^{*\alpha}$.

PROOF. (a) Let $\phi \in \tilde{S}_{RMG}$. Observe that

$$v(nJ) = Q^J v((n-1)J) \geqslant \tilde{q}(\phi) + \tilde{P}(\phi) v((n-1)J)$$
$$\geqslant [I + \dots + \tilde{P}^{n-1}(\phi)]\tilde{q}(\phi) + \tilde{P}^n(\phi) v(0).$$

Hence,

$$g^* = \lim_{n \to \infty} \frac{v(nJ)}{nJ} \geqslant \frac{\tilde{\Pi}(\phi)\tilde{q}(\phi)}{J} = \tilde{g}^*/J. \tag{4.5}$$

Next, let $f \in S_{RMG}$, and define $\phi \in \tilde{S}_R$, such that $\phi^{r,i} = f$ for all $i \in \Omega$, $r = 1, \dots, J$; observe that

$$\tilde{g}^* \geqslant \tilde{g}(\phi) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \tilde{P}^k(\phi)\tilde{q}(\phi)$$
$$= \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} P^{kJ}(f)[I + \dots + P^{J-1}(f)]q(f) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{nJ-1} P(f)^k q(f)$$
$$= J(\Pi(f)q(f)) = Jg^*$$

which together with (4.5) proves part (a).

(b) Recall that $g^* \geqslant P(f)g^*$ for any $f \in S_R$. If $\xi \in \tilde{L}(i)$, then, for each $r = 1, \ldots, J$

$$\begin{aligned} P(f^1) \ldots P(f^r) g_i^* &\leqslant g_i^* \\ &= \sum_j \tilde{P}_{ij}^{\xi} g_j^* = P(f^1) \cdots P(f^r)\left[P(f^{r+1}) \cdots P(f^J) g^*\right]_i \\ &\leqslant P(f^1) \cdots P(f^r) g_i^*. \end{aligned}$$

Hence, $P(f^1) \cdots P(f^r) g_i^* = g_i^*$. When $r = 1$, this implies $g_i^* = \sum_j P(f^1)_{ij} g_j^*$ and when $r \geqslant 2$, this implies that $[P(f^r)g^*]_j = g_j^*$ for all j such that $P(f^1) \cdots P(f^{r-1})_{ij} > 0$.

(c) Fix $v^* \in V$, and $i \in \Omega$, take $\xi = (f^1, \ldots, f^J) \in \tilde{L}(i)$ and observe from part (b) that

$$\begin{aligned} &v_i^* \geqslant q(f^1)_i - g_i^* + \left[P(f^1)v^*\right]_i. \\ &v_j^* \geqslant q(f^2)_j - g_j^* + \left[P(f^2)v^*\right]_j, \quad \text{for all } j \text{ such that } P(f^1)_{ij} > 0. \\ &\quad\vdots \\ &v_j^* \geqslant q(f^J)_j - g_j^* + \left[P(f^J)v^*\right]_j, \quad \text{for all } j \text{ such that } P(f^1) \ldots P(f^{J-1})_{ij} > 0. \end{aligned}$$

Insert the J inequalities successively into each other and conclude that

$$v_i^* \geqslant \tilde{q}_i^{\xi} + \sum_j \tilde{P}_{ij}^{\xi} v_j^* - J g_i^*, \quad \text{for all } \xi \in \tilde{L}(i)$$

where the equality sign holds for $\xi = (f^1, \ldots, f^J)$ iff

$$\begin{aligned} &b(v^*, f^1)_i = 0, \\ &b(v^*, f^r)_j = 0 \quad \text{for all } j \text{ such that } P(f^1) \cdots P(f^{r-1})_{ij} > 0;\ r = 2, \ldots, J. \end{aligned} \tag{4.6}$$

We conclude that

$$v_i^* + \tilde{g}_i^* = \max_{\xi \in \tilde{L}(i)} \left\{ \tilde{q}_i^{\xi} + \sum_j \tilde{P}_{ij}^{\xi} v_j^* \right\} = \tilde{T} v_i^* \quad \text{for all } i \in \Omega, \text{ or } v^* \in \tilde{V}.$$

Hence $V \subseteq \tilde{V}$. The dimensions of V and $\tilde{V}$ follow from theorem 5.5 in [13].

(d) Apply lemma 2.2 part (a) to the "J-step" MDP, and use the fact that $v \in \tilde{V}$ (cf. part (c)), in order to show that $\phi \in \tilde{S}_{RMG}$ iff

$$\begin{aligned} &\tilde{\phi}_{i\xi} > 0 \Rightarrow \xi \in \tilde{L}(i) \quad \text{for all } i \in \Omega, \\ &\tilde{b}(v, \phi)_i = 0 \quad \text{for all } i \in \tilde{R}(\phi). \end{aligned} \tag{4.7}$$

Use part (b), (4.6) and (4.2) in order to prove that (4.7) is equivalent to (4.3).

(e) Fix $\alpha \in \{1, \ldots, n^*\}$ and $r, t \in \{1, \ldots, d(\alpha)\}$ such that $t = r + kJ$ (modulo $d(\alpha)$) for some $k = 1, 2, \ldots$. It then follows from theorem 3.1 part (j) and (2.5) that $P(f)_{ij}^{nd(\alpha)+kJ} > 0$ for all n sufficiently large, $i \in R^{*\alpha, r}$ and $j \in r^{*\alpha, t}$. Since $\tilde{P}(\phi) = P(f)^J$, it follows that $\tilde{P}(\phi)_{ij}^{nd(\alpha)+k} > 0$, for all n sufficiently large, $i \in R^{*\alpha, r}$ and $j \in R^{*\alpha, t}$ which shows that all the states in each of the sets in (4.4) communicate with each other for $\tilde{P}(\phi)$. In addition, we observe, using theorem 3.1 part (g) that each of the sets in (4.4) is closed under $\tilde{P}(\phi)$ as well which proves that all of these sets are subchains of $\tilde{P}(\phi)$, and $\tilde{R}(\phi) \supseteq R^*$. We complete the proof of parts (e) (1) and (2), by showing the reversed inclusion $\tilde{R}(\phi) \subseteq R^*$, merely noting that for all $i \in \Omega \backslash R^*$,

$$\tilde{\Pi}(\phi)_{ii} = \lim_{n \to \infty} \tilde{P}(\phi)_{ii}^n = \lim_{n \to \infty} P(f)_{ii}^{nJ} = 0.$$

We next fix $\alpha \in \{1, \dots, n^*\}$, $t \in \{1, \dots, d(\alpha)\}$ and a state $i \in R^{*\alpha, t}$. Observe from theorem 3.1 part (j) that $R^{*\alpha, t}$ is a cyclically moving subset of $P(f)$ and use (2.4) and (2.6) in order to show

$$P(f)^n_{ii} > 0 \Rightarrow P(f)^n_{ij} > 0 \quad \text{for all } n \text{ sufficiently large, and all } j \in R^{*\alpha, t},$$

$$P(f)^n_{ii} > 0 \Rightarrow P(f)^n_{ij} = 0 \quad \text{for all } n = 1, 2, \dots \text{ and } j \notin R^{*\alpha, t}. \tag{4.8}$$

Note, using $\tilde{P}(\phi) = P(f)^J$ that (4.8) holds for $\tilde{P}(\phi)$ as well and conclude that each of the $R^{*\alpha, t}$ is a cyclically moving subset of $\tilde{P}(\phi)$, thus proving part (e) (3).

(f), (g) and (h) Fix $\phi \in \tilde{S}_{RMG}$ and let $\tilde{C}$ be a subchain of $\tilde{P}(\phi)$. Define

$$\bar{T} = \bigcup_{i \in \tilde{C}} \bigcup_{r=1}^{J} T^{r, i}(\phi)$$

(cf. (4.2)) and observe that

$$\tilde{C} = \bigcup_{i \in \tilde{C}} T^{J, i}(\phi),$$

hence

$$\tilde{C} \subseteq \bar{T}. \tag{4.9}$$

For each $j \in \bar{T}$, let $A_j = \{(r, i) \mid 1 \leqslant r \leqslant J, i \in \tilde{C} \text{ and } j \in T^{r, i}(\phi)\}$.

Next fix $v \in V$, and define $f \in S_R$ such that

$$\{k \mid f_{jk} > 0\} = \begin{cases} \bigcup_{(r, i) \in A_j} \{k \mid \phi^{r+1, i}_{jk} > 0\} & \text{for } j \in \bar{T}, \\ \{k \in L(j) \mid b(v)^k_j = 0\} & \text{for } j \notin \bar{T}. \end{cases}$$

Use part (d) in order to show that for all $i \in \Omega$: $b(v, f)_i = 0$ and $f_{ik} > 0$ only for $k \in L(i)$, hence $f \in S_{RMG}$ via lemma (2.2) part (a). Since $\bar{T}$ is closed, and the states in $\bar{T}$ communicate with each other for $P(f)$, we conclude that $\bar{T}$ is a subchain of $P(f)$. This implies using lemma 2.1 part (c) that

$$\tilde{C} \subseteq \bar{T} \subseteq R^{*\alpha} \quad (\text{for one } \alpha, 1 \leqslant \alpha \leqslant n^*) \tag{4.10}$$

which proves $\tilde{R}^* \subseteq R^*$ and hence part (f), the reversed inclusion $\tilde{R}^* \supseteq R^*$ following from part (e) (1).

Next, fix $i \in \tilde{C}$. We then have in view of (4.10) that $i \in R^{*\alpha, t}$ (for some t, $1 \leqslant t \leqslant d(\alpha)$). Use the fact that $\bar{T} \subseteq R^{*\alpha}$, and theorem 3.1 part (g) in order to show successively that

$$T^{r, i}(\phi) \subseteq R^{*\alpha, t+r} \quad \text{for } r = 1, \dots, J.$$

In particular, we obtain that

$$\begin{aligned} &\{j \mid \tilde{P}(\phi)_{ij} > 0\} = T^{J, i}(\phi) \subseteq R^{*\alpha, t+J} \text{ so that} \\ &\tilde{C} = \{j \mid \tilde{P}(\phi)^k_{ij} > 0 \text{ for some } k = 1, 2, \dots\} \subseteq \bigcup_{k=1}^{\infty} R^{*\alpha, t+kJ} \end{aligned} \tag{4.11}$$

which together with part (e) (2) proves part (g), using lemma 2.1 part (f).

Finally, a repeated application of (4.11) shows that

$$\tilde{P}(\phi)^n_{ii} > 0 \Rightarrow \tilde{P}(\phi)^n_{ij} = 0 \quad \text{for all } j \notin R^{*\alpha, t}, \text{ and all } n = 1, 2, \dots$$

which in view of (2.6) shows that each of the cyclically moving subsets of each of the policies in $\tilde{S}_{RMG}$ lies within *one* $R^{*\alpha, \beta}$. This, in combination with part (e) (3), proves part (h), using theorem 3.1 part (k). ∎

REMARK 3. It is well known from Markov Chain Theory that the chain structure of the Jth power of a single stochastic matrix $P(f)$ is related to the chain structure of $P(f)$ in the following way:

(a) the states that are transient (recurrent) for $P(f)$ are transient (recurrent) for $P^J(f)$.

(b) One obtains the subchains of $P^J(f)$ as follows: for each subchain $C^m(f)$ $(m = 1, \ldots, n(f))$, partition the collection of cyclically moving subsets $\{C^{m,t}(f) \mid t = 1, \ldots, d^m(f)\}$ (where the numbering of the c.m.s. satisfies (2.4)) into g.c.d. $\{J, d^m(f)\}$ subcollections, such that

(1) each of the subcollections contains exactly $d^m(f)/\text{g.c.d. } \{J, d^m(f)\}$ c.m.s.

(2) the rank numbers of the c.m.s. within the same subcollection differ by a multiple of g.c.d. $\{J, d^m(f)\}$ (modulo $d^m(f)$).

(c) the collection of all the c.m.s. of $P(f)$ and the one of $P^J(f)$ coincide. Parts (f), (g) and (h) of the previous theorem show that the same correspondence holds with respect to the chain structure of the set of "one-step" maximal gain policies, and the one of the set of "J-step" maximal gain policies.

Consider, for instance, the "2-step" MDP in example 1:

TABLE 2

case	J	$\tilde{n}^*$	$\tilde{R}^{*1}$	$\tilde{d}(1)$	$\tilde{R}^{*1,1}$	$\tilde{R}^{*2}$	$\tilde{d}(2)$	$\tilde{R}^{*2,1}$	$\tilde{R}^{*2,2}$
3	2	3	{1}	1	{1}	{2, 5}	1	{2, 5}	—
4	2	3	{1}	1	{1}	{2, 5}	2	{2}	{5}
3	4	3	{1}	1	{1}	{2, 5}	1	{2, 5}	—
4	4	5	{1}	1	{1}	{2}	1	{2}	—

$\tilde{R}^{*3}$	$\tilde{d}(3)$	$\tilde{R}^{*3,1}$	$\tilde{R}^{*3,2}$	$\tilde{R}^{*4}$	$\tilde{d}(4)$	$\tilde{R}^{*4,1}$	$\tilde{R}^{*5}$	$\tilde{d}(5)$	$\tilde{R}^{*5,1}$
{3, 4}	1	{3, 4}	—	—	—	—	—	—	—
{3, 4}	2	{3}	{4}	—	—	—	—	—	—
{3, 4}	1	{3, 4}	—	—	—	—	—	—	—
{4}	1	{4}	—	{5}	1	{5}	{3}	1	{3}

(Verify that $\tilde{n}^* = \sum_{\alpha=1}^{2} \text{g.c.d. } \{J, d(\alpha)\}$ and that $\tilde{d}(\alpha) = d(2)/\text{g.c.d. } \{J, d(2)\}$ for $\alpha = 2, \ldots, \tilde{n}^*$.)

Define d^* = least common multiple (l.c.m.) of $\{d(\alpha) \mid 1, \ldots, n^*\}$.

The following corollary will be needed for the analysis of the asymptotic behavior of $v(n)$:

COROLLARY 4.2. *Let $J = d^*$. Then*

(a) $\{\tilde{R}^{*\gamma} \mid \gamma = 1, \ldots, \tilde{n}^*\} = \{R^{*\alpha, t} \mid \alpha = 1, \ldots, n^*;\ t = 1, \ldots, d(\alpha)\}$.

(b) $\tilde{n}^* = \sum_{\alpha=1}^{n^*} d(\alpha)$.

(c) $\tilde{d}(\gamma) = 1$ *for all* $\gamma = 1, \ldots, \tilde{n}^*$.

5. The asymptotic behavior of $v(n)$.

In this section we study the asymptotic behavior of $v(n)$. We show that $\{v(nJ + r) - (nJ + r)g^*\}_{n=1}^{\infty}$ converges for every final reward vector $v(0)$ if and only if J is a multiple of d^*, and as a consequence that $\{v(n) - ng^*\}_{n=1}^{\infty}$ converges for every vector $v(0)$ if and only if there exists an aperiodic randomized maximal gain policy that has R^* as its set of recurrent states.

THEOREM 5.1. (a) *$\{v(n) - ng^*\}_{n=1}^{\infty}$ is bounded.*

(b) (*cf. Lanery* [7, *proposition* 7].) *If $f \in S_{RMG}$, and C is a subchain of $P(f)$ with period d, then $\lim_{n\to\infty}[v(nd + r) - (nd + r)g^*]_i$ exists for all $i \in C$, $r = 0, \ldots, d - 1$ and $v(0) \in E^N$.*

(c) $\lim_{n\to\infty}[v(nd(\alpha)+r)-(nd(\alpha)+r)g^*]_i$ *exists for all* $i \in R^{*\alpha}$, $\alpha = 1, \ldots, n^*$, $r = 1, \ldots, d(\alpha)$ *and* $v(0) \in E^N$.

(d) $\lim_{n\to\infty}[v(nd^*+r)-(nd^*+r)g^*]_i$ *exists for all* $i \in \Omega$, $r = 1, \ldots, d^*$ *and all* $v(0) \in E^N$.

PROOF. (a) cf. Brown [3, corollary 4.3] and Schweitzer and Federgruen [15, remark 1].

(b) Note that

$$v(n+1)_i \geqslant q(f)_i + P(f)v(n)_i, \qquad i \in C.$$
$$(n+1)g_i^* = g_i^* + nP(f)g_i^*, \qquad i \in C,$$

since $f \in S_{RMG}$ (cf. lemma 2.2, part (a)).

$$v_i^* = q(f)_i - g_i^* + P(f)v_i^*, \qquad i \in C, \text{ for any } v^* \in V.$$

Fix $v^* \in V$, let $e(n) = v(n) - ng^* - v^*$, and subtract the above equalities from the inequality, in order to get $e(n+1)_i \geqslant P(f)e(n)_i$, $i \in C$, and by induction

$$e(md+nd+r)_i \geqslant P(f)^{md}e(nd+r)_i, \qquad i \in C. \tag{5.1}$$

It follows from part (a) that each of the sequences $\{v(nd+r)_i - (nd+r)g_i^*\}_{n=1}^{\infty}$ and hence $\{e(nd+r)_i\}_{n=1}^{\infty}$, $i \in C$, has at least one limit point. For all $i \in C$, let x_i and y_i be two limit points of the sequence $\{e(nd+r)_i\}_{n=1}^{\infty}$. Consider (sub)sequences $\{n_k\}_{k=1}^{\infty}$ and $\{m_k\}_{k=1}^{\infty}$ of the sequence of positive integers, such that $\lim_{k\to\infty} e(n_k d + r)_i = x_i$, $i \in C$, and $\lim_{k\to\infty} e(m_k d + n_k d + r)_i = y_i$, $i \in C$. Replace in (5.1) n and m by n_k and m_k, and let k tend to infinity, in order to conclude

$$y_i \geqslant \sum_{j\in C} \bar{\pi}_{ij} x_j, \qquad i \in C, \tag{5.2}$$

where $\bar{\pi}_{ij} = \lim_{n\to\infty} P(f)_{ij}^{nd}$; $i, j \in C$. Multiply (5.2) by $\bar{\pi} \geqslant 0$ to get $\bar{\pi}y \geqslant \bar{\pi}x$. Since x and y are arbitrary limit points, we have the reversed inequality $\bar{\pi}x \geqslant \bar{\pi}y$ as well, hence $\bar{\pi}x = \bar{\pi}y$. As a consequence, (5.2) becomes

$$y_i \geqslant \sum_{j\in C} \bar{\pi}_{ij} y_j, \qquad i \in C.$$

Multiply these inequalities by $\bar{\pi} \geqslant 0$, and note $\bar{\pi}_{ii} > 0$, for all $i \in C$ (cf. (2.6)), to conclude that

$$y_i = [\bar{\pi}y]_i, \qquad i \in C.$$

Thus,

$$y_i = \bar{\pi}y_i = \bar{\pi}x_i = x_i \quad \text{for all } i \in C$$

which proves that $\{e(nd+r)_i\}_{n=1}^{\infty}$ has exactly one limit point, for all $i \in C$.

(c) Take f^* as in lemma (2.2) part (d), and apply part (b), using theorem 3.1 part (a).

(d) It suffices to prove that $\lim_{n\to\infty}[Q^{nd^*}v(0) - nd^*g^*]$ exists for all $v(0)$, because then

$$\lim_{n\to\infty}\left[v(nd^*+r)-(nd^*+r)g^*\right] = \lim_{n\to\infty}\left[Q^{nd^*}v(r) - nd^*g^*\right] - rg^*$$

will also exist for all $v(0)$ and all $r = 1, \ldots, d^*$.

Define $\tilde{Q} = Q^{d^*}$ and consider the d^*-step MDP, as described in §4. Note $v(nd^*) - nd^*g^* = \tilde{Q}^n v(0) - n\tilde{g}^*$ (cf. theorem 4.1 part (a)). Fix $v(0)$ and define

$$x_i = \lim_{n\to\infty} \inf\left[\tilde{Q}^n v(0) - n\tilde{g}^*\right]_i; \quad X_i = \lim_{n\to\infty} \sup\left[\tilde{Q}^n v(0) - n\tilde{g}^*\right]_i, \qquad i \in \Omega.$$

From part (a), it follows that $-\infty < x_i \leqslant X_i < \infty$ for all i. Observe, using (2.12) that for all n sufficiently large

$$\left[\tilde{Q}^{n+1}v(0) - (n+1)\tilde{g}^*\right]_i = \left[\tilde{T}\tilde{Q}^n v(0) - (n+1)\tilde{g}^*\right]_i = \left[\tilde{T}\left[\tilde{Q}^n v(0) - n\tilde{g}^*\right] - \tilde{g}^*\right]_i$$

$$= \max_{\xi \in \tilde{L}(i)} \left\{ \tilde{q}_i^\xi - \tilde{g}_i^* + \sum_j \tilde{P}_{ij}^\xi \left[\tilde{Q}^n v(0) - n\tilde{g}^*\right]_j \right\}, \qquad i \in \Omega. \quad (5.3)$$

Fix $i \in \Omega$, take (sub)sequences $\{n_k\}_{k=1}^\infty$ (with $\lim_{k\to\infty} n_k = \infty$) such that

$$\lim_{k\to\infty} \left[\tilde{\mathrm{Q}}^{n_k} v(0) - n_k \tilde{g}^*\right]$$

exists and $\lim_{k\to\infty} [\tilde{Q}^{n_k+1}v(0) - (n_k+1)\tilde{g}^*]_i = x_i$ (or X_i resp.). Replace n by n_k in (5.3), and let k tend to infinity in order to conclude

$$X_i \leqslant \max_{\xi \in \tilde{L}(i)} \left[\tilde{q}_i^\xi - \tilde{g}_i^* + \sum_j \tilde{P}_{ij}^\xi X_j\right], \qquad i \in \Omega. \quad (5.4)$$

$$x_i \geqslant \max_{\xi \in \tilde{L}(i)} \left[\tilde{q}_i^\xi - \tilde{g}_i^* + \sum_j \tilde{P}_{ij}^\xi x_j\right], \qquad i \in \Omega. \quad (5.5)$$

If ϕ achieves the N maxima in (5.4), we have

$$\tilde{q}(\phi) - \tilde{g}^* + \tilde{P}(\phi)x \leqslant x \leqslant X \leqslant \tilde{q}(\phi) - \tilde{g}^* + \tilde{P}(\phi)X \quad (5.6)$$

or

$$0 \leqslant X - x \leqslant \tilde{P}(\phi)(X - x),$$

whence we get, by iterating this inequality

$$0 \leqslant X - x \leqslant \tilde{\Pi}(\phi)(X - x).$$

We complete the proof of showing $X - x = 0$ by demonstrating that $(X - x)_i = 0$ for all $i \in \tilde{R}(\phi)$. Multiply the right inequality in (5.6) by $\tilde{\Pi}(\phi) \geqslant 0$, noting that ϕ has support on $\times_{i\in\Omega} \tilde{L}(i)$, in order to get

$$0 \leqslant \tilde{\Pi}(\phi)\left[\tilde{q}(\phi) - \tilde{g}^*\right] = \tilde{g}(\phi) - \tilde{g}^* \leqslant 0,$$

where the last inequality follows from (2.8). Hence $\phi \in \tilde{S}_{RMG}$ and $\tilde{R}(\phi) \subseteq \tilde{R}^* = R^*$ (cf. theorem 4.1 part (f)) which proves $(X - x)_i = 0$, $i \in \tilde{R}(\phi)$, since part (c) shows that $(X - x)_i = 0$ for all $i \in R^*$. ∎

We next show that the sequences $\{v(nJ + r) - (nJ + r)g^*\}_{n=1}^\infty$ do not converge for all final reward vectors $v(0)$, unless J is a multiple of d^*. However, we first need the following lemma.

LEMMA 5.2. *Define $\tilde{Q} = Q^{d^*}$, and consider the corresponding "d^*-step" MDP. Let $\tilde{T}$, $\tilde{V}$ be defined as in §4, and fix $v \in V$.*

(a) *For all $\tilde{v} \in \tilde{V}$, we have $\tilde{v} = v + x$, where there are $\tilde{n}^*$ constants $\{y^{\alpha,t} \mid \alpha = 1, \ldots, n^*; t = 1, \ldots, d(\alpha)\}$ with the convention that the superscript t in $y^{\alpha,t}$ is taken modulo $d(\alpha)$, such that for all $\alpha \in \{1, \ldots, n^*\}$, and $t \in \{1, \ldots, d(\alpha)\}$:*

$$x_i = y^{\alpha,t} \quad \text{for all } i \in R^{*\alpha,t}, \quad (5.7)$$

$$(T^m\tilde{v})_i = v_i + mg_i^* + y^{\alpha,t+m} \quad \text{for all } i \in R^{*\alpha,t}; \qquad m = 0, 1, 2, \ldots. \quad (5.8)$$

(b) *$\tilde{v} \in \tilde{V}$ can be chosen such that all the $y^{\alpha,t}$ are distinct.*

PROOF. (a) Observe, using theorem 4.1 part (c) that $v \in \tilde{V}$, and use theorem 5.1 of Schweitzer and Federgruen [14] in order to show (5.7).

Next, take $f \in S^*_{RMG}$ and observe, using lemma (2.2) part (a) that

$$T^m \tilde{v} \geqslant q(f) + P(f) T^{m-1} \tilde{v}, \qquad m = 1, \ldots, d^*. \tag{5.9}$$

Using the fact that $\tilde{v} \in \tilde{V}$ and inserting the d^* inequalities in (5.9) successively into each other, we obtain

$$\tilde{v} + d^* g^* = \tilde{T}\tilde{v} \geqslant T^{d^*}\tilde{v} \geqslant \left[I + \cdots + P(f)^{d^*-1} \right] q(f) + P(f)^{d^*} \tilde{v}. \tag{5.10}$$

By multiplying (5.10) with $\Pi(f) \geqslant 0$, we conclude strict equality for all components $i \in R^*$. It next follows from (5.9) that

$$T^{d^*}\tilde{v}_i = \left[q(f) + P(f) T^{d^*-1}\tilde{v} \right]_i \quad \text{for all } i \in R^*,$$

and more generally that

$$\left[T^k \tilde{v} \right]_i = \left[q(f) + P(f) T^{k-1} \tilde{v} \right]_i \quad \text{for all } k = 1, \ldots, d^* \text{ and}$$

$$i \in \left\{ i \mid P(f)_{ji}^{d^*-k} > 0 \text{ for some } j \in R^* \right\} = R^*, \tag{5.11}$$

where the last equality follows from $R(f) = R^*$.

We next prove (5.8) for $m = 0, \ldots, d^*$. It then follows that (5.8) holds for any value of m, since for all $n = 1, 2, \ldots$ and $m = 1, \ldots, d^*$

$$T^{nd^*+m}\tilde{v}_i = T^m (T^{nd^*}\tilde{v})_i = T^m(\tilde{v} + nd^* g^*)_i = nd^* g_i^* + T^m \tilde{v}_i$$

$$= nd^* g_i^* + v_i + m g_i^* + y^{\alpha, t+m} = v_i + (nd^* + m) g_i^* + y^{\alpha, t+nd^*+m}$$

for all $i \in R^{*\alpha, t}$ ($1 \leqslant \alpha \leqslant n^*$; $1 \leqslant t \leqslant d(\alpha)$). First observe that (5.8) holds for $m = 0$. Next assume it holds for $m = k$, with $0 \leqslant k < d^*$. It then follows that (5.8) holds for $m = k + 1$, as well since, using (5.11), and theorem 3.1 part (g)

$$(T^{k+1}\tilde{v})_i = \left[q(f) + P(f) T^k \tilde{v} \right]_i$$

$$= q(f)_i + \sum_{j \in R^{*\alpha, t+1}} P(f)_{ij} \left\{ v_j + k g_j^* + y^{\alpha, t+k+1} \right\}$$

$$= 0 + v_i + (k+1) g_i^* + y^{\alpha, t+k+1}.$$

(b) It follows from theorem 5.5 in Schweitzer and Federgruen [14] that the $\tilde{n}^*$ parameters $\{ y^{\alpha, t} \mid \alpha = 1, \ldots, n^*;\ t = 1, \ldots, d(\alpha) \}$ may be chosen independently over some (finite) region in $E^{\tilde{n}^*}$.

THEOREM 5.3. (a) *Fix* $\alpha \in \{1, \ldots, n^*\}$, $i \in R^{*\alpha}$, $J \geqslant 1$, *and* $r \in \{0, \ldots, J-1\}$; $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r) g_i^*$ *exists for all* $v(0)$ *only if* J *is a multiple of* $d_i = d(\alpha)$.
(b) *Fix* $J \geqslant 1$ *and* $r \in \{0, \ldots, J-1\}$; $\lim_{n\to\infty} v(nJ + r) - (nJ + r) g^*$ *exists for all* $v(0) \in E^N$ *only if* J *is a multiple of* d^*.

PROOF. (a) Fix $v \in V$, and choose $\tilde{v} \in \tilde{V}$ as in part (b) of the previous lemma. Pick λ large enough that $Q^n(\tilde{v} + \lambda g^*) = T^n(\tilde{v} + \lambda g^*)$, for $n = 1, 2, \ldots$ (cf. (2.11)). Finally, let $i \in R^{*\alpha, t}$ ($1 \leqslant t \leqslant d(\alpha)$). Observe that $\tilde{v} + \lambda g^* \in \tilde{V}$, and apply lemma 5.2 part (a) in order to show

$$Q^{nJ+r}(\tilde{v} + \lambda g^*)_i = T^{nJ+r}(\tilde{v} + \lambda g^*)_i = \lambda g_i^* + v_i + (nJ + r) g_i^* + y^{\alpha, t+nJ+r}.$$

Hence,

$$Q^{nJ+r}(\tilde{v} + \lambda g^*)_i - (nJ + r) g_i^* = v_i + \lambda g_i^* + y^{\alpha, t+nJ+r}.$$

Since $\lim_{n\to\infty} Q^{nJ+r}(\tilde{v} + \lambda g^*)_i - (nJ + r)g_i^*$ exists and since the $y^{\alpha, t}$ ($\alpha = 1, \ldots, n^*$; $t = 1, \ldots, d(\alpha)$) are chosen to be distinct, we must have $t + nJ + r$ (modulo $d(\alpha)$) $= \gamma$ (say) for all n large enough, which implies that J is a multiple of $d(\alpha)$.

(b) Since $\lim_{n\to\infty}[v(nJ + r) - (nJ + r)g^*]_i$ exists for all $i \in R^*$ and $v(0) \in E^N$, it follows from part (a) that J must be a multiple of the $d(\alpha)$ ($\alpha = 1, \ldots, n^*$) hence J is a multiple of d^*.

Combining theorem 5.1 parts (c) and (d), with theorem 5.3, we obtain our main result.

THEOREM 5.4. (a) *Fix* $\alpha \in \{1, \ldots, n^*\}$, $i \in R^{*\alpha}$, *and two integers* J *and* r. *Then* $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r)g_i^*$ *exists for all* $v(0) \in E^N$ *if and only if* J *is a multiple of* $d(\alpha) = d_i$.

(b) $\lim_{n\to\infty} v(nJ + r) - (nJ + r)g^*$ *exists for all* $v(0) \in E^N$ *if and only if* J *is a multiple of* d^*.

REMARK 4. The following conditions are equivalent statements of the necessary and sufficient condition for the convergence of $\{v(n) - ng^*\}_{n=1}^{\infty}$, for all $v(0) \in E^N$.

(I) $d^* = 1$.

(II) There exists an aperiodic *randomized* maximal gain policy f, *with* $R(f) = R^*$.

(III) Each state $i \in R^*$ lies within an aperiodic subchain of some *randomized* maximal gain policy.

(IV) For each $\alpha \in \{1, \ldots, n^*\}$ there exists a *randomized* maximal gain policy which has an aperiodic subchain within $R^{*\alpha}$.

(Observe that (I)$\Rightarrow$(II) as a result of theorem 3.1 part (a), (II)$\Rightarrow$(III) and (III) $\Rightarrow$(IV) are immediate, while (IV)$\Rightarrow$(I) is immediate from (3.1)).

We note that in (II), (III) and (IV) the adjective "randomized" cannot be replaced by "pure"; in fact, the modification of example 1, case 1 where $K(5) = \{1, 2\}$ shows that $d^* = 1$ can occur, with *all* of the *pure* policies being periodic.

Moreover, example 1, case 3 and 4 show that the addition "with $R(f) = R^*$" in (II) is indispensable: f^6 is an aperiodic maximal gain policy, however, with $R(f^6) \subset R^*$.

Finally example 1, case 3, with $d^* = 2$, shows that $\lim_{n\to\infty} v(n) - ng^*$ fails to exist for some $v(0) \in E^5$ (take $v(0) = [2q_5^2\ q_5^2\ 0\ 0\ q_5^2]$, observe that $v(2n + 1) = [2q_5^2\ 0\ q_5^2\ q_5^2\ 0]$ and $v(2n) = [2q_5^2\ q_5^2\ 0\ 0\ q_5^2]$. Note that $v(0) \in \tilde{V} \setminus V$ and cf. theorem 5.3).

THEOREM 5.5. *The following conditions are sufficient for the existence of* $\lim_{n\to\infty}[v(n) - ng^*]$ *for all* $v(0) \in E^N$.

(I) *All of the transition probabilities are strictly positive.*

$$P_{ij}^k > 0, \quad \text{for all } i, j \in \Omega, \text{ and } k \in K(i)$$

(*cf. Bellman* [2], *Brown* [3]).

(II) *For all* $v(0) \in E^N$, *there exists an aperiodic* $f \in S_P$, *and an integer* n_0, *such that*

$$v(n + 1) = q(f) + P(f)v(n), \quad \text{for all } n \geqslant n_0$$

(*cf. Morton* [9]).

(III) *There exists a state* s *and an integer* $\nu \geqslant 1$, *such that*

$$P(f^1) \cdots P(f^\nu)_{is} > 0 \quad \text{for all } f^1, f^2, \ldots, f^\nu \in S_P;\ i \in \Omega$$

(*cf. White* [17]).

(IV) *Every* $f \in S_P$ *is aperiodic* (*cf. Schweitzer* [12] *and* [13]).

(V) *Every* $f \in S_{PMG}$ *is aperiodic* (*cf. Schweitzer* [12] *and* [13]).

(VI) *For each* $i \in R^*$, *there exists a pure maximal gain policy* f, *such that state* i *is recurrent and aperiodic for* $P(f)$.

(VII) *Every pure maximal gain policy has a unichained tpm, and at least one of them is aperiodic.*

PROOF. (I)$\Rightarrow$(III)$\Rightarrow$(IV)$\Rightarrow$(V)$\Rightarrow$(VI) where the last implication follows from lemma (2.1) part (a).

(VI)$\Rightarrow d_i = 1$ for all $i \in R^* \Rightarrow d^* = d(\alpha) = 1$ for all $\alpha = 1, \ldots, n^*$ (cf. theorem 3.1. part (c)). The sufficiency of (II) follows from the fact that after n_0 iterations the policy space may be reduced to $S_P^{new} = \{f\}$ which satisfies (IV).

(VII)$\Rightarrow n^* = 1$, since the subchains of any two tpm's must intersect, and in addition $d^* = d(1) = 1$ as a consequence of theorem 3.2. ∎

We have seen that for arbitrary $J \geqslant 1$, and some fixed $v(0)$ the sequences $\{v(nJ + r)_i - (nJ + r)g_i^*\}_{n=1}^{\infty}$ may fail to converge for some (or all) $i \in \Omega$ and for some (or all) $r = \{0, 1, \ldots, J-1\}$.

However, the various sequences interdepend as far as their asymptotic behaviour is concerned.

We conclude this section by exhibiting the various ways in which this interdependence occurs. However we first need the following lemma.

LEMMA 5.6. *Fix $f \in S_{RMG}$. Then*

$$\lim_{n\to\infty} \left[v(n+1)_i - q(f)_i - P(f)v(n)_i \right] = 0, \quad \textit{for all } i \in R(f).$$

PROOF. Use the fact that for all $i \in \Omega$, $f_{ik} > 0$ only for $k \in L(i)$ (cf. lemma 2.2 part (a)(1)) in order to show that

$$v(n+1) - (n+1)g^* \geqslant q(f) - g^* + P(f)\left[v(n) - ng^* \right]. \tag{5.12}$$

By multiplying (5.12) with $\Pi(f)$, we obtain

$$\Pi(f)(v(n+1) - (n+1)g^*) \geqslant \Pi(f)(v(n) - ng^*).$$

Observing from theorem 5.1 part (a) that $\Pi(f)(v(n) - ng^*)$ is bounded in n, we conclude the existence of $L = \lim_{n\to\infty} \Pi(f)(v(n) - ng^*)$. Define

$$\delta(n) = v(n+1) - q(f) - P(f)v(n)$$

and note that $\delta(n) \geqslant 0$ for all n (cf. (1.1)). Then

$$\begin{aligned} \lim_{n\to\infty} \Pi(f)\delta(n) &= \lim_{n\to\infty} \{\Pi(f)[v(n+1) - (n+1)g^*] \\ &\qquad - \Pi(f)(q(f) - g^*) - \Pi(f)(v(n) - ng^*)\} \\ &= L - L = 0, \end{aligned}$$

which proves the lemma using $\delta(n) \geqslant 0$ and the fact that $\Pi(f) \geqslant 0$ with $\Pi(f)_{jj} > 0$ for all $j \in R(f)$.

THEOREM 5.7. (a) *Fix $\alpha \in \{1, \ldots, n^*\}$; $\lim_{n\to\infty} v(n)_i - ng_i^*$ exists either for all $i \in R^{*\alpha}$ or for none of them.*

(b) *Fix $J \geqslant 1$ and $i \in R^{*\alpha, t}$ ($1 \leqslant \alpha \leqslant n^*$; $1 \leqslant t \leqslant d(\alpha)$). Assume $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r)g_i^*$ exists for some integer r. Then*

$$\lim_{n\to\infty} v(nJ + r + s)_j - (nJ + r + s)\, g_j^* \quad \textit{exists for all } j \in \bigcup_{k=1}^{\infty} R^{*\alpha,\, t+kJ-s} \qquad (s = 1, 2, \ldots). \tag{5.13}$$

(c) *Fix $J \geqslant 1$, and $\alpha \in \{1, \ldots, n^*\}$. $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r)g_i^*$ exists for all $r = 1, \ldots, J$, either for all $i \in R^{*\alpha}$ or for none of them.*

(d) *Fix* $J \geqslant 1$, $r_0 \in \{1, \ldots, J\}$ *and* $\alpha \in \{1, \ldots, n^*\}$. *If* $\lim_{n\to\infty} v(nJ + r_0)_i - (nJ + r_0)g_i^*$ *exists for all* $i \in R^{*\alpha}$, *then* $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r)g_i^*$ *exists for all* $i \in R^{*\alpha}$ *and all* $r = 1, 2, \ldots$.

(e) *Fix* $i \in \Omega$. *Assume* $\lim_{n\to\infty} v(nJ^1 + r)_i - (nJ^1 + r)g_i^*$, *and* $\lim_{n\to\infty} v(nJ^2 + s)_i - (nJ^2 + s)g_i^*$ *exist for all* $r \in \{1, \ldots, J^1\}$ *and* $s \in \{1, \ldots, J^2\}$. *Let* $J^3 = \text{g.c.d.}\{J^1, J^2\}$. *Then* $\lim_{n\to\infty} v(nJ^3 + t)_i - (nJ^3 + t)g_i^*$ *exists for all* $t = 1, \ldots, J^3$, *and hence, if in addition* $i \in R^{*\alpha}$ (*for some* $1 \leqslant \alpha \leqslant n^*$) *then* $\lim_{n\to\infty} v(nJ^3 + t)_j - (nJ^3 + t)g_j^*$ *exists for all* $t = 1, \ldots, J^3$ *and all* $j \in R^{*\alpha}$.

(f) *Fix* $i \in R^{*\alpha}$ $(1 \leqslant \alpha \leqslant n^*)$ *and* $J \geqslant 1$. *Assume* $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r)g_i^*$ *exists for all* $r = 1, \ldots, J$. *Let* $\hat{J} = \text{g.c.d.}\{J, d(\alpha)\}$. *Then* $\lim_{n\to\infty} v(n\hat{J} + s)_j - (n\hat{J} + s)g_j^*$ *exists for all* $s = 1, \ldots, \hat{J}$ *and all* $j \in R^{*\alpha}$.

(g) *Fix* $J \geqslant 1$. *If* $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r)g_i^*$ *exists for all* $i \in R^*$ *and some* $r \in \{1, \ldots, J\}$, *then* $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r)g_i^*$ *exists for all* $i \in \Omega$ *and all* $r = 1, 2, \ldots$.

PROOF. (a) Assume $\lim_{n\to\infty} v(n)_t - ng_t^*$ exists for some $t \in R^{*\alpha}$. Define

$$x_i = \lim_{n\to\infty} \inf \left[v(n) - ng^* \right]_i; \qquad X_i = \lim_{n\to\infty} \sup \left[v(n) - ng^* \right]_i,$$

and observe that $-\infty < x_i \leqslant X_i < \infty$ as a result of theorem 5.1 part (a). Fix $i \in R^{*\alpha}$, pick $\epsilon > 0$ and apply lemma 5.6 with $f^* \in S^*_{RMG}$ in order to show that there exists an integer $n(\epsilon)$, such that for all $n > n(\epsilon)$

$$q(f^*)_i - g^*{}_i + P(f^*)\left[v(n) - ng^*\right]_i \leqslant v(n+1)_i - (n+1)\, g_i^*$$
$$\leqslant q(f^*)_i - g_i^* + P(f^*)\left[v(n) - ng^*\right]_i + \epsilon. \qquad (5.14)$$

Take (sub)sequences $\{n_k\}_{k=1}^{\infty}$ (with $\lim_{k\to\infty} n_k = \infty$) such that $\lim_{k\to\infty}[v(n_k) - n_k g^*]$ exists and

$$\lim_{k\to\infty} v(n_k + 1)_i - (n_k + 1)\, g_i^* = x_i \text{ (or } X_i).$$

Replace n by n_k in (5.14) and let k tend to infinity in order to conclude

$$q(f^*)_i - g_i^* + P(f^*)x_i \leqslant x_i \leqslant X_i \leqslant q(f^*)_i - g_i^* + P(f^*)X_i + \epsilon,$$

or

$$0 \leqslant X_i - x_i \leqslant P(f^*)(X - x)_i, \quad \text{for all } i \in R^{*\alpha},$$

whence we get by iterating this inequality

$$0 \leqslant X_i - x_i \leqslant \langle \pi^\alpha(f^*), X - x \rangle, \qquad i \in R^{*\alpha}. \qquad (5.15)$$

Multiply this inequality by $\Pi(f^*) \geqslant 0$ in order to conclude strict equality on the right of (5.15), thus proving $X_i - x_i = X_t - x_t = 0$, for all $i \in R^{*\alpha}$.

(b) Without loss of generality, we take $r = 0$. Define $\tilde{Q} = Q^J$ and consider the J-step MDP, as defined in §4. Let f^* be defined as in lemma 2.2 part (d), and let $\overline{R}(s) = \bigcup_{k=1}^{\infty} R^{*\alpha, t+kJ-s}$ for $s = 1, 2, \ldots$. Observe that $v(nJ)_i - nJg^*{}_i = [\tilde{Q}^n v(0)]_i - n\tilde{g}_i^*$ (cf. theorem 4.1 part (a)). Apply part (a) of this theorem to the J-step MDP, and use theorem 4.1 part (g) in order to obtain that $v(nJ)_i - nJg_i^*$ exists for all $i \in \overline{R}(0)$, thus proving (5.13) for $s = 0$. Assume (5.13) holds for $s = S$. Note, using theorem 3.1 part (f) that for all $i \in \overline{R}(S + 1)$, $P(f^*)_{ij} > 0$ only for $j \in \overline{R}(S)$. It then follows from lemma 5.6 that for all $i \in \overline{R}(S + 1)$

$$\lim_{n\to\infty} v(nJ + S + 1)_i - (nJ + S + 1)\, g_i^* = q(f^*)_i - g_i^* + \sum_{j \in \overline{R}(S)} P(f^*)_{ij} x_j$$

where

$$x_i = \lim_{n\to\infty} v(nJ + S)_i - (nJ + S)g_i^*, \quad \text{for all } i \in \bar{R}(S),$$

which proves part (b) by complete induction.

(c) Assume $\lim_{n\to\infty} v(nJ + r)_j - (nJ + r)g_j^*$ exists for *all* $r = 1, \ldots, J$ and $j \in R^{*\alpha, t}$ $(1 \leqslant t \leqslant d(\alpha))$. Take $i \in R^{*\alpha, u}$ $(1 \leqslant u \leqslant d(\alpha))$ and $s \in \{1, \ldots, J\}$. Then $\lim_{n\to\infty} v(nJ + s)_i - (nJ + s)g_i^*$ exists as a result of part (b).

(d) Take $i \in R^{*\alpha, t}$ $(1 \leqslant t \leqslant d(\alpha))$ and $r \in \{1, \ldots, J\}$; $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r)g_i^*$ exists as a result of part (b).

(e) Let $p^1 = J^1/J^3$ and $p^2 = J^2/J^3$. Fix $t \in \{1, \ldots, J^3\}$, and define $a(n) = v(nJ^3 + t)_i - (nJ^3 + t)g_i^*$. Observe that $A(m) = \lim_{n\to\infty} a(np^2 + m)$ exists for all $m = 1, \ldots, p^2$, just as $A = \lim_{n\to\infty} a(np^1)$ exists. Observe that there exist two integers $\alpha, \beta \geqslant 1$ such that $\alpha p^1 - \beta p^2 = 1$, as a consequence of p^1 and p^2 being relatively prime. Since

$$A(m) = \lim_{k\to\infty} a\left[(kp^1 + \beta m)p^2 + m\right] = \lim_{k\to\infty} a\left[(kp^2 + \alpha m)p^1\right] = A,$$

for all $m = 1, \ldots, p^2$, it follows that $\lim_{n\to\infty} a(n)$ exists, thus proving the first assertion, whereas the second one follows immediately from part (c).

(f) Use part (e) with $J^1 = J$ and $J^2 = d(\alpha)$ (cf. theorem 5.1 part (c)).

(g) It follows from part (c) that $\lim_{n\to\infty} v(nJ + r)_i - (nJ + r)g_i^*$ exists for all $i \in R^*$ and *all* $r \in \{1, \ldots, J\}$ whereas convergence on $\Omega\backslash R^*$ is deduced, using the proof of theorem 5.1 part (d). ∎

REMARK 5. The following statements illustrate the degree of interdependence with respect to the asymptotic behavior of the N sequences $\{v(n)_i - ng_i^*\}$ $(i \in \Omega)$, and may be proved using the above theorem, merely verifying all possible combinations.

(a) $\lim_{n\to\infty} v(n)_i - ng_i^*$ cannot exist for all values of i, but one (cf. Schweitzer [13, theorem 1 part (3)]).

(b) If $\lim_{n\to\infty} v(n)_i - ng_i^*$ exists for all values of i except two, then these two special states comprise one $R^{*\alpha}$, with $d(\alpha) = 2$. Moreover, for every randomized maximal gain policy, these two states either form a periodic subchain, or are both transient.

(c) If $\lim_{n\to\infty} v(n) - ng^*_i$ exists for all values of i except three, then either the three states comprise one $R^{*\alpha}$ with $d(\alpha) = 2$ or 3, or else two of them comprise one $R^{*\alpha}$ with $d(\alpha) = 2$, and the third one lies in $\Omega - R^*$, having positive probability to reach $R^{*\alpha}$.

The generalization of theorem 5.4 for the case of *one* fixed $v(0)$ is

THEOREM 5.8. (a) *Fix $v(0)$, and $\alpha \in \{1, \ldots, n^*\}$. There exists an integer $J^{0\alpha} \geqslant 1$, dependent upon $v(0)$, such that $\lim_{n\to\infty}[v(nJ + r) - (nJ + r)g^*]_i$ exists for all $i \in R^{*\alpha}$ and some r if and only if the integer $J \geqslant 1$ is a multiple of $J^{0\alpha}$. If this condition is met, the limit exists for all r. The integer $d(\alpha)$ is a multiple of $J^{0\alpha}$. If $d(\alpha) \geqslant 2$, then there exist choices of $v(0)$ such that $J^{0\alpha} < d(\alpha)$ can occur.*

(b) *Fix $v(0)$ and define the integer $J^0 = \text{l.c.m.}\{J^{0\alpha} \mid 1 \leqslant \alpha \leqslant n^*\}$ which depends upon $v(0)$. Then $\lim_{n\to\infty} v(nJ + r) - (nJ + r)g^*$ exists for some r if and only if the integer $J \geqslant 1$ is a multiple of J^0. If this condition is met, the limit exists for all r. The integer d^* is a multiple of J^0. If $d^* \geqslant 2$, then there exist choices of $v(0)$ such that $J^0 < d^*$ can occur.*

PROOF. (a) Let

$$J^{0\alpha} = \text{g.c.d.}\left\{J \geqslant 1 \mid \lim_{n\to\infty}\left[v(nJ + r)_i - (nJ + r)g_i^*\right] \text{ exists for all } i \in R^{*\alpha}\right\}. \tag{5.16}$$

Observe that $J^{0\alpha}$ can be obtained as the g.c.d. of a finite number of integers and apply theorem 5.7 part (e) to conclude that $J^{0\alpha}$ belongs to the set to the right of (5.16), thus proving the first assertion. The second and third assertions follow from theorem 5.7 part (d) and theorem 5.1 part (c), whereas the last one may be verified by choosing

$$v(0) = v + tg^* \quad \text{with } v \in V \text{ and } t \text{ sufficiently large that}$$

$$Q^n(v(0)) = T^n(v(0)) \text{ for } n = 1, 2, \ldots, \tag{5.17}$$

so $J^{0\alpha} = 1$.

(b) Observe from part (a) that $\lim_{n\to\infty} [v(nJ + r)_i - (nJ + r)g_i^*]$ exists for all $i \in R^*$, and some r if and only if J is a multiple of J^0, and apply part (g) of theorem 5.7 to verify the first two assertions. The third assertion follows from theorem 5.1 part (d) whereas the existence of $v(0)$ with $J^0 = 1$ may be verified by choosing $v(0)$ as in (5.17). ∎

Note 1. More specifically the following argument was used in (VII-64) and (VII-75).

$$\max_{k=1,\ldots,K} \max_{i,j \in A_k} \{f_i - f_j\} = \max_{1 \leq i,j \leq n} \{f_i - f_j\}$$

where f is an n-vector and the $\{A_k, k = 1, \ldots, K\}$ constitute a partition of $\{1, \ldots, n\}$.

The assertions (VII-64) and (VII-75) are repeatedly used in the remainder of the proof.

References

[1] Bather, J. (1973). Optimal Decision Procedures for Finite Markov Chains, Part I, II. *Advances in Appl. Probability* **5** 328–339, 521–540.

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

[3] Brown, B. (1965). On the Iterative Method of Dynamic Programming on a Finite State Space Discrete Time Markov Process. *Ann. Math. Statist.* **36** 1279–1285.

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

[5] Hordijk, A., Schweitzer, P. J. and Tijms, H. (1975). The Asymptotic Behaviour of the Minimal Total Expected Cost for the Denumerable State Markov Decision Model. *J. Appl. Probability* **12** 298–305.

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

[7] Lanery, E. (1967). Etude Asymptotique des Systèmes Markoviens à Commande. *Rev. Inf. Rech. Op.* **3** 3–56.

[8] Lembersky, M. R. (1974). On Maximal Rewards and ϵ-Optimal Policies in Continuous Time Markov Decision Chains. *Ann. Statist.* **2** 159–169.

[9] Morton, T. (1971). On The Asymptotic Convergence Rate of Cost Differences for Markovian Decision Processes. *Operations Res.* **19** 244–248.

[10] Odoni, A. (1969). On Finding the Maximal Gain for Markov Decision Processes. *Operations Res.* **17** 857–860.

[11] Romanovskii, V. (1970). *Discrete Markov Chains.* Wolters-Noordhoff, Groningen.

[12] Schweitzer, P. J. (1965). Perturbation Theory and Markovian Decision Processes. Ph.D. dissertation, M.I.T. Operations Research Center Report 15.

[13] ———. (May 1968). A Turnpike Theorem for Undiscounted Markovian Decision Processes. Presented at ORSA/TIMS national meeting San Francisco, California.

[14] ——— and Federgruen, A. (to appear). Functional Equations of Undiscounted Markov Renewal Programming. *Math. of Oper. Res.*

[15] ——— and ———. (1977). Geometric Convergence of Value-Iteration in Multichain Markov Decision Problems. Math Center Report BW 80/77.

[16] Tijms, H. (1975). On Dynamic Programming with Arbitrary State Space, Compact Action Space and the Average Return Criterion. Math. Center report BW 55/75.

[17] White, D. (1963). Dynamic Programming, Markov Chains, and the Method of Successive Approximations. *J. Math. Anal. Appl.* **6** 373–376.

I.B.M. THOMAS J. WATSON RESEARCH CENTER, YORKTOWN HEIGHTS, NEW YORK 10598 (and) GRADUATE SCHOOL OF MANAGEMENT, UNIVERSITY OF ROCHESTER, ROCHESTER, NEW YORK 14627

FOUNDATION MATHEMATISCH CENTRUM, 2e BOERHAAVESTRAAT 49, AMSTERDAM 1005, THE NETHERLANDS