stichting
mathematisch
centrum

AFDELING MATHEMATISCHE BESLISKUNDE BW 80/77 OKTOBER
(DEPARTMENT OF OPERATIONS RESEARCH)

P.J. SCHWEITZER & A. FEDERGRUEN

GEOMETRIC CONVERGENCE OF VALUE-ITERATION IN MULTICHAIN MARKOV RENEWAL PROGRAMMING

2e boerhaavestraat 49 amsterdam

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

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

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

Geometric convergence of value-iteration in multichain Markov renewal programming

by

P.J. Schweitzer * & A. Federgruen **

ABSTRACT

This paper considers undiscounted Markov Decision Problems. With no restriction (on either the periodicity - or chain structure of the problem) we show that the value iteration method for finding maximal gain policies, exhibits a geometric rate of convergence, whenever convergence occurs.
In addition, we study the behaviour of the value-iteration operator; we give bounds for the number of steps needed for contraction, describe the ultimate behaviour of the convergence factor and give conditions for the existence of a *uniform* convergence rate.

KEY WORDS & PHRASES: *Markov Decision Problems; average cost criterion; value-iteration method; geometric convergence; convergence factor; existence of a uniform convergence rate.*

* I.B.M. Thomas J. Watson Research Center, Yorktown Heights, N.Y. 10598, U.S.A.

** Mathematisch Centrum, Amsterdam, Netherlands.

1. INTRODUCTION AND SUMMARY

The value-iteration equations for undiscounted Markov Decision Processes (MDP's) were first studied by BELLMAN [2] and HOWARD [9]:

$$(1.1.) \qquad v(n+1)_i = Qv(n)_i, \qquad i = 1,\dots N$$

where the value-iteration operator Q is defined by:

$$(1.2.) \qquad Qx_i = \max_{k\in K(i)} \{q_i^k + \sum_{j=1}^{N} P_{ij}^k v(n)\}_j, \qquad i = 1,\dots,N;\ x \in E^N.$$

and with $v(0)$ a given N-vector. $K(i)$ denotes the finite set of alternatives in state i, q_i^k the one-step expected reward and P_{ij}^k the transition probability to state j when alternative $k \in K(i)$ is chosen in state i ($i = 1,\dots,N$). For all $n = 1,2,\dots$ and $i \in \Omega = \{1,\dots,N\}$, $v(n)_i$ denotes the maximal total expected reward for a planning horizon of n epochs obtained when ending up at state j.

BROWN [3] showed that $\{v(n)-ng^*\}_{n=1}^{\infty}$ is uniformly bounded in n, provided g^* is taken as the maximal gain rate vector. In [18] we proved the existence of an integer J such that

$$(1.3.) \qquad u(r) = \lim_{n\to\infty} [v(nJ+r) - (nJ+r)g^*]$$

exists for *all* $v(0) \in E^N$ and $r = 0,\dots,J-1$. (Previous proofs in [3] and [10] are both incorrect or incomplete.)

In general $\{v(n)-ng^*\}_{n=1}^{\infty}$ may fail to converge for arbitrary $v(0)$ if some of the transition probability matrices (<u>tpm's</u>) are periodic i.e. $J \geq 2$ can occur. Sufficient conditions for the convergence of $\{v(n)-ng^*\}_{n=1}^{\infty}$ for *all* $v(0) \in E^N$, were obtained by BATHER [1], LANERY [10], SCHWEITZER [14,15] and WHITE [22], while the necessary and sufficient condition was recently obtained in [18]. While the result in [18] settles the issue if one demands existence of $\lim_{n\to\infty} \{v(n)-ng^*\}_{n=1}^{\infty}$ for every $v(0) \in E^N$, it should be noted that $\{v(n)-ng^*\}_{n=1}^{\infty}$ always converges for $v(0)$ belonging to some non-empty closed set $W \subseteq E^N$ (cf. lemma 2.2).

In this paper we return to the issue of the *rate* of convergence. Our main result (th.4.2) is the fact that if $\lim_{n\to\infty} v(n)-ng^*$ exists, then the approach to the limit is *geometric*. Consequently this result shows that the *value-iteration method* for locating maximal gain policies (cf. [12],[14] and [22]) exhibits a geometric rate of convergence. This result is of particular importance to the case N >> 1 where this value-iteration method is the only feasible one for finding maximal gain policies.

This generalization of White's result (cf. [22] to the general multichain case is remarkable since the property of geometric convergence holds in spite of the fact that the operator Q is *never* a contraction mapping or a J-step contraction mapping for any J = 1,2,... (cf. DENARDO [4] and [7]) with respect to any norm on E^N. Note e.g. that for all $x \in E^N$ and scalars c:

$$(1.4.) \qquad Q(x+c\underline{1}) = Q\,x + c\underline{1}, \qquad \text{with } \underline{1} \text{ the N-vector of ones.}$$

In addition, and even more remarkably, the Q-operator, does not need to be (J-step) contracting (for any $J \geq 1$) with respect to the following pseudo-norm either (cf. [1]):

$$(1.5.) \qquad \|x\|_d = x_{max} - x_{min}, \qquad x \in E^N$$

with $x_{max} = \max_i x_i$ and $x_{min} = \min_i x_i$, the use of which is suggested by the very property (1.4.) (cf. BATHER [1]).

Indeed although we find a convergence rate (or *ultimate* average contraction factor per step) which is strictly bounded away from one on W, the average contraction factor per step may initially be very close to one; and in general there does not exist an integer $n \geq 1$ such that the *n-step* contraction factor is strictly bounded away from 1 (cf. section 7).

One should point out that the geometric convergence result holds for all $v(0) \in W$, with *no* restrictions imposed on e.g. - the chain - and periodicity structure. In addition if v(0) is such that (1.3.) holds with $J \geq 2$ the same th.4.2 applied to a related "J-step" decision process shows that the approach to the limit in (1.2.) will be geometric for each r = 0,...,J-1 as well.

In section 2 we give the notation and preliminaries. In section 3 we study the evolution of the Q-operator. The geometric convergence result is obtained in section 4. In section 5 we give some additional properties for MDP's satisfying condition (H1) to be stated below; in particular, we show that the number of steps needed for contraction is bounded by a quadratic function in N. In section 6 we characterize the ultimate behaviour of the Q-operator and of the average contraction factor per step. In section 7, finally, we derive for MDP's satisfying (H1) the necessary and sufficient condition for the existence of a uniform n-step contraction factor (for some $n \geq 1$) - i.e. a n-step contraction factor which is strictly bounded away from one on W.

We refer to [7] for some necessary and some sufficient conditions for the Q-operator to be contracting with respect to the $\|\cdot\|_d$ norm.

2. NOTATION AND PRELIMINARIES

A (stationary) randomized policy is a tableau $[f_{ik}]$ satisfying $f_{ik} \geq 0$ and $\sum_{k \in K(i)} f_{ik} = 1$ (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 $f(i) = k$, where k denotes the single alternative used in state i. Associated with each $f \in S_R$, are the N-component "reward" vector q(f) and the N x N matrix P(f):

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

Note that P(f) is a stochastic matrix, for any $f \in S_R$, and define the stochastic matrix $\Pi(f)$ as the Cesaro limit of the sequence $\{P^n(f)\}_{n=1}^{\infty}$. Define the *maximal-gain rate vector* g^*:

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

DERMAN [5] proved that there exists a *pure* policy that achieves the N suprema in (2.2) simultaneously. In addition Howard's Policy Iteration Algorithm (cf. [9]) showed that the quantities $a_i^k = \sum_{j=1}^N P_{ij}^k g_j^* - g_i^*$, $i \in \Omega$, $k \in K(i)$ satisfy:

$$(2.3) \qquad \max_{k \in K(i)} a_i^k = 0 \qquad , i = 1,\dots,N,$$

as well as the existence of vectors v^* satisfying the optimality equation:

$$(2.4) \qquad v_i^* = \max_{k \in L(i)} \{q_i^k - g_i^* + \sum_j P_{ij}^k v_j^*\}, \qquad i = 1,\dots,N, \text{ where}$$

$$L(i) = \{k \in K(i) \mid a_i^k = 0\}, \qquad i = 1,\dots,N.$$

Accordingly define S_{PMG} and S_{RMG} as the set of pure and randomized maximal-gain policies i.e.

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

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

Let $R(f) = \{j \in \Omega \mid \Pi(f)_{jj} > 0\}$ i.e. $R(f)$ is the set of recurrent states for $P(f)$, and define $R^* = \bigcup_{f \in S_{RMG}} R(f)$.

In th.3.2. of [17] we proved that

$$(2.5) \qquad R^* = \bigcup_{f \in S_{PMG}} R(f).$$

and that there exists $f \in S_{RMG}$ with $R(f) = R^*$. Let V denote the non-empty solution set to the optimality equation (2.4). Observe that if $v \in V$ then $v + c_1\underline{1} + c_2 g^* \in V$ for all scalars c_1, c_2. For any $v \in E^N$, define

$$(2.6) \qquad b(v)_i^k = q_i^k - g_i^* + \sum_{j=1}^N P_{ij}^k v_j - v_i; \qquad i \in \Omega, k \in K(i)$$

and

and

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

Note that $\max_{k\in L(i)} b(v)_i^k = 0$ for every $i \in \Omega$, if and only if $v \in V$. As a consequence we define for any $v \in V$:

(2.7) $$L(i,v) = \{t \in L(i) \mid b(v)_i^t = \max_{k\in L(i)} b(v)_i^k = 0\}.$$

In th.3.1. part (e) of [17] we established the following characterization of S_{RMG}:

(2.8) Fix $v \in V$. Let $f \in S_R$; then $f \in S_{RMG}$ if and only if: $f_{ik} > 0$ implies $k \in L(i,v)$ for all $i \in R(f)$ and $k \in L(i)$ for all $i \in \Omega\backslash R(f)$.

In addition to the pseudo-norm $\| \ \|_d$ (cf.(1-5)) we will use the norm $\|x\|_\infty = \max_i |x_i|$. Note that

(2.9) $$x_{min} \le 0 \le x_{max} \Rightarrow \|x\|_\infty \le \|x\|_d; \quad x \in E^N.$$

Finally, define for $x \in E^N$:

(2.10) $$x^+ = \begin{cases} \min\{x_i \mid x_i > 0, i \in \Omega\} & \text{if } x_{max} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$x^- = \begin{cases} \max\{x_i \mid x_i > 0, i \in \Omega\} & \text{if } x_{min} < 0 \\ 0 & \text{otherwise} \end{cases}$$

Lemma 2.1. below enumerates a number of elementary properties of the Q-operator that will be needed in the remainder. First, let Q^n denote the n-fold application of the operator:

$$Q^n x = Q(Q^{n-1}x); \quad n = 1,2,... \text{ and } x \in E^N, \text{ with } Q^0x = x$$

and define for all $x \in W$, $L(x)=\lim_{n\to\infty} Q^n x - ng^*$:

LEMMA 2.1.

(a) $(x-y)_{min} \le (Qx-Qy)_{min} \le (Qx-Qy)_{max} \le (x-y)_{max}; x,y \in E^N$
(b) $\|Qx-Qy\|_d \le \|x-y\|_d$; $\|Qx-Qy\|_\infty \le \|x-y\|_\infty; x,y \in E^N$
(c) *If* $x,y \in W$ *then for* $n = 0,1,\ldots$:

$$(Q^n x-Q^n y)_{min} \le (L(x)-L(y))_{min} \le (L(x)-L(y))_{max} \le (Q^n x-Q^n y)_{max}$$

and

$$\|L(x)-L(y)\|_d \le \|Q^n x-Q^n y\|_d;\ \|L(x)-L(y)\|_\infty \le \|Q^n x-Q^n y\|_\infty.$$

(d) $L(x)$ *is a Lipschitz continuous function on* W.
(e) W *is closed and unbounded.*
(f) *If* $x \in W$, *then* $Q^m x \in W$ *for all* $m = 1,2,\ldots$ *and* $L(Q^m x)=L(x) + mg^*$.
(g) *Suppose* $(Qx-Qy)_{max} = (x-y)_{max}$; *state* r *satisfies* $(Qx-Qy)_r = (Qy-Qy)_{max}$ *and alternative* $k \in K(r)$ *achieves* $(Qx)_r$, i.e. $(Qx)_r = q_r^k + \sum_j P_{rj}^k x_j$.
Then $(Qy)_r = q_r^k + \sum_j P_{rj}^k y_j$ *as well, and* $P_{rs}^k > 0$ *only if state s satisfies* $(x-y)_s = (x-y)_{max}$.
(h) *Similary, if* $(Qx-Qy)_r = (Qx-Qy)_{min} = (x-y)_{min}$ *for some* $r \in \Omega$ *and* $k \in K(r)$ *achieves* $(Qy)_r$, i.e. $(Qy)_r = q_r^k + \sum_j P_{rj}^k y_j$ *then* k *achieves* $(Qx)_r$ *as well and* $P_{rs}^k > 0$ *only if* $(x-y)_s = (x-y)_{min}$.

PROOF: The proof of part (a) is easy and may be found in lemma 2.1 of [1]; part (b) follows from part (a). A repeated application of (a) shows that for all $n,m \ge 0$: $(Q^n x-Q^n y)_{min} \le [(Q^{n+m}x-(n+m)g^*) - (Q^{n+m}y-(n+m)g^*)]_{min} \le [(Q^{n+m}x-(n+m)g^*) - (Q^{n+m}y-(n+m)g^*)]_{max} \le (Q^n x-Q^n y)_{max}$.
Next, the first assertion of part (c) follows by letting m tend to infinity, whereas the second assertion and part (d) are an immediate consequence of the first one.
Next, consider a sequence $\{x^\alpha\}_{\alpha=1}^\infty$ with $x^\alpha \in W$, $\alpha = 1,2,\ldots$ and $\lim_{\alpha\to\infty} x^\alpha = x^*$. Pick $\varepsilon > 0$ and x^α such that $\|x^\alpha-x^*\|_\infty < \varepsilon/3$.

Since $x^{\alpha} \in W$, there is some $n_0(\varepsilon) \geq 1$ such that for all $n,m \geq n_0(\varepsilon)$:

$$\|(Q^n x^{\alpha} - ng^*) - (Q^m x^{\alpha} - mg^*)\|_{\infty} < \varepsilon/3.$$

Hence, for all $n,m \geq n_0(\varepsilon)$:

$$\begin{aligned}\|(Q^n x^* - ng^*) - (Q^m x^* - mg^*)\|_{\infty} &\leq \|Q^n x^* - Q^n x^{\alpha}\|_{\infty} + \\ \|(Q^n x^{\alpha} - ng^*) - (Q^m x^{\alpha} - mg^*)\|_{\infty} &+ \|Q^m x^* - Q^m x^{\alpha}\|_{\infty} \\ \leq 2\|x^* - x^{\alpha}\|_{\infty} + \varepsilon/3 = \varepsilon,\end{aligned}$$

the last inequality following from part (a). Hence, by Cauchy's convergence criterion, $\lim_{n\to\infty} Q^n x^* - ng^*$ exists, which proves that W is closed, whereas W is unbounded in view of $x \in W$ implying $x + c\underline{1} \in W$ for any scalar c, with $L(x+c\underline{1}) = L(x) + c\underline{1}$, thus proving part (e).

(f): follows from $\lim_{n\to\infty} Q^n(Q^m x) - ng^* = \lim_{n\to\infty} \{Q^{n+m}x - (n+m)g^*\} + mg^* = L(x) + mg^*$.

The proofs of part (g) and (h) are easy, and may also be found in BATHER [1], lemma 2.2. □

In addition to the Q-operator defined by (1.2), we introduce:

(2.11) $$Tx_i = \max_{k\in L(i)} \{q_i^k + \sum_{j=1}^{N} P_{ij}^k x_j\}, \quad i \in \Omega;\ x \in E^N.$$

We let T^n denote the n-fold application of the T-operator and in analogy to W and $L(x), x \in W$ we define:

$$\tilde{W} = \{x \in E^N \mid \lim_{n\to\infty} T^n x - ng^* \text{ exists}\} \text{ and for all } x \in \tilde{W},$$

$$\tilde{L}(x) = \lim_{n\to\infty} T^n x - ng^*.$$

Observe that the T operator is the value-iteration operator associated with a related MDP in which the policy space is restricted to $X_{i=1}^{N} L(i)$. As a consequence it has *all* of the properties of the Q-operator as exhibited in the previous lemma.

The following lemma shows that the Q-operator reduces to the T operator in at least two ways, and that the latter has a number of additional

properties which induce that the sequence $\{T^n x\}_{n=1}^{\infty}$ has a more regular behaviour than $\{Q^n x\}_{n=1}^{\infty}$.
First define:

(2.12) $$e(n,x) = Q^n x - ng^* - L(x), \quad x \in W, \quad n \geq 0.$$
$$\tilde{e}(n,x) = T^n x - ng^* - \tilde{L}(x), \quad x \in \tilde{W}, \quad n \geq 0.$$

By definition, $\lim_{n\to\infty} e(n,x) = 0$ for $x \in W$ and $\lim_{n\to\infty} \tilde{e}(n,x) = 0$ for $x \in \tilde{W}$:

LEMMA 2.2.

(a) $T(x+cg^*) = Tx + cg^*$ *for any scalar* c. *If* $x \in \tilde{W}$, *then for any scalar* c, $x + tg^* \in \tilde{W}$ *and* $\tilde{L}(x+cg^*) = \tilde{L}(x) + cg^*$.

(b) *For any* $v \in V$, $T^n v = v + ng^*$. *Also* $V \subseteq \tilde{W}$ *and* $\tilde{L}(v) = v$ *for any* $v \in V$.

(c) *For any* $n \geq 1$ *and* $i = 1,\ldots N$.

(2.13) $$e(n+1,x)_i = \max_{k\in K(i)} [na_i^k + b(L(x))_i^k + \sum_{j=1}^{N} P_{ij}^k e(n,x)_j], \quad x \in W$$

(2.14) $$\tilde{e}(n+1,x)_i = \max_{k\in L(i)} [b(\tilde{L}(x))_i^k + \sum_j P_{ij}^k \tilde{e}(n,x)_j], x \in \tilde{W}.$$

(d) *For each* $x \in E^N$ *there exists an integer* $n_0(x)$ *such that* $Q^{n_0+m}x = T^m(Q^{n_0}x)$ *for* $m = 1,2,\ldots$. *Also if* $x \in W$, *then* $Q^{n_0}x \in \tilde{W}$ *with* $\tilde{L}(Q^{n_0}x) = L(x)+n_0 g^*$.

(e) *For all* $x \in W$:

$$e(n+1,x)_{\min} \geq e(n,x)_{\min}; \qquad n = 0,1,\ldots$$

$$e(n+1,x)_{\max} \leq \begin{cases} e(n,x)_{\max}; & n > n_0(x) \\ \max_{i,k} b(L(x))_i^k + e(n,x)_{\max}; & n \leq n_0(x) \end{cases}$$

Hence, for all $x \in \tilde{W}$ *and* $n = 0,1,\ldots$:

(2.15) $\quad \tilde{e}(n,x)_{min} \le \tilde{e}(n+1,x)_{min} \le 0 \le \tilde{e}(n+1,x)_{max} \le \tilde{e}(n,x)_{max}$, *and*

$$\|\tilde{e}(n+1,x)\|_d \le \|\tilde{e}(n;x)\|_d;\ \|\tilde{e}(n+1,x)\|_\infty \le \|\tilde{e}(n,x)\|_\infty$$

(f) *For each* $x \in E^N$ *there exists a scalar* $t_0(x)$ *such that*
$Q^n(x+tg^*) = T^n(x+tg^*)$ *for* $n = 0,1,2,\dots$ *and* $t \ge t_0(x)$.
Hence if $v \in V$ *then* $v + tg^* \in W$ *if* t *large enough i.e.* W *is non-empty.*

(g) *For any* $x \in W, L(x) \in V$ *and for any* $x \in \tilde{W}, \tilde{L}(x) \in V$.

(h) $\tilde{W}\backslash V = \{x \in \tilde{W} \mid \|x-\tilde{L}(x)\|_d > 0\}$.

<u>PROOF</u>.

(a) Immediate from the definition of L(i).

(b) For $v \in V$, $Tv = v + g^*$ follows from (2.4). By induction, we obtain $T^n v = v + ng^*$.

(c) Part (c) follows straightforward from the definitions (2.3), (2.6) and (2.12).

(d) The fact that for large n, the Q-operator only uses alternatives in L(i) was proved in th.4.4 of [3] (cf. also remark 1).
Next, $\lim_{m\to\infty} T^m(Q^n 0x) - mg^* = \lim_{m\to\infty} \{Q^{m+n}0x-(m+n_0)g^*\}+n_0g^* = L(x)+n_0g^*$.

(e) Since by (2.7) and (2.13), $e(n+1,x)_i \ge \sum_j P^k_{ij}\, e(n,x)_j$ for $k \in L(i,L(x))$ we have $e(n+1,x)_{min} \ge e(n,x)_{min}$ for all $x \in W$. Next by (2.3):

$$e(n+1,x)_i \le \max_{k\in K(i)} \{b(L(x))^k_i + \textstyle\sum_j P^k_{ij}\, e(n,x)_j\},\ i \in \Omega \text{ so}$$

$$e(n+1,x)_{max} \le \max_{i,k} b(L(x))^k_i + e(n,x)_{max};\ n=0,1,\dots$$

Since part (d) shows that for all $n > n_0(x)$ the maximum in (2.13) is attained by an alternative in L(i), for all $i \in \Omega$, we obtain the sharper bound $e(n+1,x)_{max} \le e(n,x)_{max}$ for all $n > n_0(x)$ in view of (2.7). Next, the outer inequalities in (2.15) follow immediately from the above, whereas the inner ones are due to $\{\tilde{e}(n,x)\quad\}^\infty_{n=0}$ and $\{\tilde{e}(n,x)_{max}\}^\infty_{n=0}$ being monotonically non-decreasing and non-increasing to $\lim_{n\to\infty}\tilde{e}(n,x)_{min} = \lim_{n\to\infty}\tilde{e}(n,x)_{max} = 0$.

(f) Fix $v \in V$. By repeating the proof of part (e) with repect to $\tilde{e}(n,x) = T^n x - ng^* - v$, for any $x \in E^N$, one shows that $\{T^n x - ng^*\}^\infty_{n=1}$

is bounded for all $x \in E^N$ (cf. also BROWN [3] and remark 1).

$$Q(T^n x+tg^*)_i = \max_{k\in K(i)} \{(t+n)a_i^k + (t+n)g_i^* + q_i^k + \sum_j P_{ij}^k[T^n x-ng^*]_j\}, i \in \Omega$$

it follows that there exists a scalar $t_0(x)$ such that for all $t \geq t_0(x)$ only alternatives in L(i) achieve the maximum, for all n = 0,1,.... Hence the first assertion of part (f) trivially holds for n = 0 and proceeding by complete induction, assume it holds for some integer n. Then $Q^{n+1}(x+tg^*) = Q[T^n(x+tg^*)] = Q[T^n x+tg^*] = T[T^n x+tg^*] = T^{n+1}(x+tg^*)$ for all $t \geq t_0(x)$. Finally if $v \in V$ and $t \geq t_0(v)$ then $Q^n(v+tg^*) - ng^* = T^n v + tg^* - ng^* = v + tg^*$ for all n = 0,1... (cf. part (b)) which proves $v + tg^* \in W$ for all $t \geq t_0(v)$.

(g) Letting n tend to infinity in (2.14) and recalling $\lim_{n\to\infty} \tilde{e}(n,x)=0$ one observes that for $x \in \tilde{W}$, $\max_{k\in L(i)} b(\tilde{L}(x))_i^k=0$; hence $\tilde{L}(x) \in V$. Since $L(x)= \tilde{L}(Q^{n_0}x) - n_0 g^*$ for any $x \in W$ (cf. part (e)) it follows that $L(x) \in V$ for any $x \in W$.

(h) Let $x \in \tilde{W}$. If $x \in V$ then $\|x-\tilde{L}(x)\|_d = 0$ follows form part (b). Conversely if $\|x-\tilde{L}(x)\|_d = 0$ then $x = \tilde{L}(x) + c\underline{1}$ for some scalar c; so $x \in V$ in view of part (g). □

3. THE EVOLUTION OF THE Q OPERATOR.

Convergence of $\{Q^n x-ng^*\}_{n=1}^{\infty}$ occurs in three phases. During the first phase the Q operator still uses alternatives in K(i)-L(i). Lemma 2.2 part (d) shows that for any $x \in E^N$ after finitely many steps namely for $n \geq n_0(x)$, alternatives in L(i) achieve the maximum in (2.13) or in other words *Q reduces to T*. (In fact the proof of this part of the lemma shows that from a certain point on, *only* alternatives in L(i) achieve the maxima).
Next, lemma 2.2 part (e) shows that the distance between $[Q^n x-ng^*]$ and its limit $L(x)$ as measured e.g. by the $\| \ \|_\infty$ is guaranteed to be monotonically non-increasing after these first $n_0(x)$ steps. This is why we say that the first $n_0(x)$ iterations constitute the *first* phase of the convergence process during which the behaviour of either $\| e(n,x)\|_d$ or $\|e(n,x)\|_\infty$ may be very irregular.

Observe that the first phase is non-existing when K(i)=L(i) for all $i \in \Omega$ as is e.g. the case when $g^* = \langle g^* \rangle \underline{1}$, i.e. when the maximal gain rate is independent of the initial state of the system.
While for $n \geq n_0(x)$ the Q-operator reduces to the T-operator, for still larger n and $x \in W$ due to $\lim_{n\to\infty} e(n,x)=0$ the maximum in (2.13) can only be achieved by alternatives for which $b(L(x))_i^k=0$ i.e. alternatives that belong to $L(i,L(x))$ (cf. (2.7)).
Hence for very large n (say for $n \geq n_1(x)$) we get the behaviour:

(3.1) $$e(n+1,x) = U(L(x))\; e(n,x), \qquad x \in W$$

where for any $v \in V$ the U(v)-operator is defined by:

(3.2) $$[U(v)y]_i = \max_{k \in L(i,v)} [\textstyle\sum_j P_{ij}^k y_j], \qquad i = 1,\ldots,N.$$

Observe that the U(v)-operator is a value-iteration operator with zero rewards. Since the associated maximal gain rate vector is $\underline{0}$ i.e. has identical components, it has all of the properties of the T-operator.
In addition it distinguishes itself by the following special (*positive homogenity*) feature:

(3.3) $$U(v)[ax] = a\, U(v)x, \qquad x \in E^N \text{ and for any scalar } a \geq 0$$

as well as by:

$$x_{max} \geq [U(v)x]_{max} \geq [U(v)x]_{min} \geq x_{min}.$$

Note that there are only a finite number of *distinct* U(v)-operators, since there are only finitely many subset of $X_i\, L(i)$.
For any $v \in V$, define:

(3.4) $$\delta(v) = \begin{cases} \infty \;, & \text{if } b(v)_i^k = 0 \text{ for all } i \in \Omega,\ k \in L(i) \\ \min\{-b(v)_i^k \mid i \in \Omega,\ k \in L(i), \text{ such that } b(v)_i^k < 0\}, & \\ \text{otherwise} & \end{cases}$$

Note that for all $x \in W$, the reduction to the $U(L(x))$-operator occurs at the very last when $\|e(n,x)\|_d$ drops below the $\delta(L(x))$-level.

We will say that the *second* phase of the convergence process starts at the $n_0(x)+1$-th iteration and terminates at the $n_1(x)$-th iteration. It is followed by the *third* phase from there on. In the following section we will show that in the second and third phase $\|e(n,x)\|_\infty$ decreases to zero at a geometric rate of convergence for all $x \in W$. Whereas the contraction factor per step initially depends upon the starting point x and may be very close to unity, the *ultimate* convergence rate or average contraction factor per step is determined by the behaviour of the U(v)-operator in the third phase and will be shown to be *uniform* i.e. strictly bounded away from one, on W.

The remainder of this section is devoted to a description of the first phase as well as to a preliminary characterization of the U(v)-operators in the third phase.

We first observe that (2.13) may be rewritten as:

$$(3.5) \qquad e(n+1,x)_i = \max_{k \in K(i)} \{b(L(Q^n x))_i^k + \textstyle\sum_j P_{ij}^k\, e(n,x)_j\}, i \in \Omega$$

since $na_i^k + b(L(x))_i^k = b(L(x)+ng^*)_i^k = b(L(Q^n x))_i^k$ the last equality following from lemma 2.1 part (f). Define:

$$(3.6) \qquad \Psi(n,x) = \max_{i \in \Omega, k \in K(i)} b(L(Q^n x))_i^k; \qquad x \in W.$$

The next theorem shows that $\{\psi(n,x)\}_{n=1}^\infty$ decreases in at least a linear way with n, so it reduces in a finite number of steps to 0, after which the non-increasing of $\|e(n,x)\|_d$ is guaranteed. Hence convergence is lexicographic in the sense that first $\{\psi(n,x)\}_{n=1}^\infty \downarrow 0$ and next $\{\|e(n,x)\|_d\}_{n=1}^\infty \downarrow 0$.

<u>THEOREM 3.1</u>. *Let* $x \in W$.

(a) $\psi(n,x) \geq 0$; $n = 0,1\ldots$. *If* $K(i) = L(i)$ *for all* i, *then* $\psi(n,x) = 0$ *for all* $n = 0,1\ldots$.

(b) $\psi(n+1,x) \leq \psi(n,x)$; *if* $\psi(n+1,x) > 0$ *then* $\psi(n+1,x) \leq \psi(n,x) + \Delta$ where

$$(3.7) \qquad \Delta = \begin{cases} \infty, & \text{if } K(i) = L(i), \quad i \in \Omega \\ \max\{a_i^k \mid a_i^k < 0;\ i \in \Omega, k \in K(i)\}, & \text{otherwise} \end{cases}$$

(c) *There exists an integer* $n_0'(x) \le \frac{\Psi(0,x)}{|\Delta|}$ *with* $\Psi(n,x)=0$ *for* $n \ge n_0'(x)$

Also $\|e(n+1,x)\|_d \le e(n,x)\|_d$ *for* $n > n_0'$.

<u>PROOF</u>.

(a) $\psi(n,x) \ge \max_{i\in\Omega,k\in L(i)} b(L(Q^n x))_i^k = 0$ since $L(Q^n x) \in V$ (cf. lemma 2.2 part (g)) while the equality sign holds if $K(i)=L(i)$ for all $i \in \Omega$.

(b) $\psi(n+1,x) = \max_{i\in\Omega,k\in K(i)}\{(n+1)a_i^k + b(L(x))_i^k\} \le$

$$\max_{i\in\Omega,k\in K(i)}\{na_i^k + b(L(x))_i^k\} = \psi(n,x)$$

Assume $\psi(n+1,x) > 0$. Then $\psi(n+1,x) = a_i^k + b(L(Q^n x))_i^k$ for some $i \in \Omega$, and $k \notin L(i)$ since $b(L(Q^n x))_i^k \le 0$ and $a_i^k = 0$ for $k \in L(i)$. Hence, $a_i^k \le \Delta$ and $\psi(n+1,x) \le \psi(n,x) + \Delta$.

(c) The existence of $n_0'(x) \le \frac{\psi(0,x)}{|\Delta|}$ follows immediately from part (b). Next, assume $\psi(n,x) = 0$ and use (2.13) to obtain:

$$e(n+1,x)_i \le \max_{k\in K(i)}\{na_i^k + b(L(x))_i^k\} + \max_{k\in K(i)}\{\textstyle\sum_j P_{ij}^k e(n,x)_j\}$$
$$\le \psi(n,x) + e(n,x)_{max} = e(n,x)_{max}$$

Hence, $e(n+1,x)_{max} \le e(n,x)_{max}$, whereas $e(n+1,x)_{min} \ge e(n,x)_{min}$ was shown in lemma 2.2 part (e). Since $\psi(n,x) = 0$ for $n \ge n_0'(x)$, we conclude that $\|e(n,x)\|_d$ is non-increasing for $n \ge n_0'$. □

Part (c) of the previous theorem shows that both $n_0(x)$ and $n_0'(x)$ are bounds on the number of iterations before $\{\|e(n,x)\|_d\}_{n=1}^{\infty}$ starts to be monotonically non-increasing. The following example will show that:

(a) the behaviour of $\|e(n,x)\|_d$ (or $\|e(n,x)\|_\infty$) may be very irregular during the first phase: in this particular example, $\|e(n,x)\|_d$ first decreases, then increases during a number of steps that is of the order of N

(b) both $n_0(x)$ and $n_0'(x)$ as defined in lemma 2.2. and th.3.1, may be very large and are not uniformly bounded in $x \in W$

(c) the convergence of $\{\psi(n,x)\}_{n=1}^{\infty}$ to 0 is exactly *linear*, i.e. $\psi(n+1,x) = \psi(n,x) + \Delta$ for all $n < n_0'(x)$

(d) both cases $n_0(x) > n_0'(x)$ and $n_0(x) < n_0'(x)$ may occur.

EXAMPLE 1:

i	k	P_{i1}^k	P_{i2}^k	P_{i3}^k	P_{i4}^k	P_{i5}^k	....	P_{iN}^k	P_{iN+1}^k	q_i^k	
1	1	1								1	
2	1	1								1	
2	2			1						0	
3	1				1					2(N-4)	
4						1				2(N-5)	$g^* =$
											(1,1,0,...,0)
N-3											
N-2								1		2	
N-1								$\frac{1}{2}$	$\frac{1}{2}$	0	
N								0	1	0	

($P_{i\,i+1}^1 = 1$ for $3 \le i \le N-2$; $q_i^1 = 2(N-i-1)$ for $3 \le i \le N-2$.
Take $x = [0,1,0,A,A,\ldots,A+\ell,A]$.
Since $L(2) = \{1\}$, for n large we have $(Q^n x)_2 - ng_2^* = 0$, hence $L(x)_1 = L(x)_2 = 0$. Moreover $L(x)_i = \sum_{r=i}^{N-1} 2(N-r-1)+A = (N-i)(N-i-1)+A$ for $i \ge 3$.
Let $\ell = 0$:

$$\|e(0,x)\|_d = e(o,x)_{max} - e(o,x)_{min} = 1+(N-4)(N-3)+A$$
$$\|e(1,x)\|_d = e(1,x)_2 - e(1,x)_3 = 0-2(N-4)+(N-4)(N-3)$$

Using $\|e(n,x)\|_d - \|e(n-1,x)\|_d = \{e(n,x)_2 - e(n-1,x)_2\} - \{e(n,x)_3 - e(n-1,x)_3\}$:

$$\|e(n,x)\|_d - \|e(n-1,x)\|_d = \begin{cases} (2(N-4)+A-1)-2(N-3) = A+1, & \text{for } n=2 \\ (2(N-n-2)-1)-2(N-n-3) = 1, & 2<n\le N-3 \\ -1 & \text{for } N-3 \le n \le A+(N-3)(N-4) \end{cases}$$

and

$$\|e(n,x)\|_d = 0 \text{ for } n > A + (N-3)(N-4).$$

$$\Delta = a_2^2 = -1;\ b(L(x))_i^1 = 0 \text{ for all } i;\ b(L(x))_2^2 = A+(N-4)(N-3)-1$$

hence

$$\psi(n,x)= \begin{cases} A+(N-4)(N-3)-(n+1) & \text{for } n<A+(N-4)(N-3) \\ 0 & \text{for } n>A+(N-4)(N-3) \end{cases}$$

and conclude that $\psi(n+1,x) = \psi(n,x) - \Delta$ for $n < n_0'(x)$.

Finally note that since the quantities $b(L(x))_i^k$ and $\psi(n,x)$ are independent of $\ell, n_0(x) > n_1(x)$ occurs when $\ell >> 0$ and $n_0(x) < n_0'(x)$ when $\ell << 0$.

REMARK 1. Fix $v^* \in V$. Let $\bar{e}(n,x) = Q^n x - ng^* - v^*$ for any $x \in E^N$, and $\bar{\psi}(n,x) = \max_{i\in\Omega, k\in K(i)} \{na_i^k + b(v^*)_i^k\}$. An examination of the proof of th.3.1 with $e(n,x)$ and $\psi(n,x)$ replaced by $\bar{e}(n,x)$ and $\bar{\psi}(n,x)$, shows that:

(a) $\{Q^n - ng^*\}_{n=1}^{\infty}$ is bounded in n, for all $x \in E^N$

Next it follows from (a) and (2.15) with $e(n,x)$ and $L(x)$ replaced by $\bar{e}(n,x)$ and v^*, that

(b) for n large enough, the Q-operator uses only alternatives in $L(i)$.

These results were already obtained in BROWN [3], who employed limiting results from the discounted case.

Lemma 3.2 below gives some preliminary properties of the $U(v)$-operator (as appearing in the third phase) and concludes this section:

LEMMA 3.2.

(a) *Fix* $v \in V$. *If* $\|y-v\|_d < \delta(v)$ *then*

$$T^n y - (ng^*+v) = T^n y - T^n v = U(v)^n (y-v):\ n = 0,1,2,\ldots$$

(b) *Take* $x \in \tilde{W}$ *with* $\|x-\tilde{L}(x)\|_d < \delta(\tilde{L}(x))$. *Then for any* $\lambda \in [0,1]$, *the vector* $x(\lambda) = (1-\lambda)\tilde{L}(x) + \lambda x$ *satisfies* $x(\lambda) \in \tilde{W}$ *and* $\tilde{L}(x(\lambda)) = \tilde{L}(x)$.

(c) *If* $v \in V$ *and the vector* p *satisfies* $\|p\|_d = 1$ *and* $\lim_{n\to\infty} U(v)^n p = 0$ *then for* $0 \le \lambda < \delta(v)$, $v + \lambda p \in \tilde{W}$ *and* $\tilde{L}(v+\lambda p) = \tilde{L}(v) = v$.

PROOF. We first observe that, by lemma 2.2 (b), $T^n v = ng^* + v$, and $T^n v \in V$, for all $n \ge 1$.

(a) $Ty_i - Tv_i = \max_{k\in L(i)} \{q_i^k + \sum_j P_{ij}^k y_j - (v_i + g_i^*)\} = \max_{k\in L(i)} \{b(v)_i^k + \sum_j P_{ij}^k (y_j - v_j)\}$

Let $k(i)$ achieve this maximum. Then applying lemma 2.1, part (a) to T:

$(y-v)_{min} \le (Ty-Tv)_{min} \le b(v)_i^{k(i)} + \sum_j P_{ij}^{k(i)}(y_j-v_j) \le b(v)_i^{k(i)} + (y-v)_{max}$, for all $i \in \Omega$. Hence $0 \le -b(v)_i^{k(i)} \le \|y-v\|_d < \delta(v)$, or $b(v)_i^{k(i)} = 0$ for $i = 1,\dots,N$ (cf. (3.4)). This proves part (a) for $n = 1$. Next, observe that by $T^n v = v + ng^*, b(v)_i^k = b(T^n v)_i^k$ for all i and $k \in L(i)$. Hence, for all $n, \delta(v) = \delta(T^n v)$ and the $U(v)$-operator and the $U(T^n v)$-operator coincide. Now assume that the assertion holds for one value of n. This implies using lemma 2.1 part (b): $\|T^n y-T^n v\|_d \le \|y-v\|_d < \delta(v) = \delta(T^n v)$, and invoking the induction assumption: $T^{n+1}y - T^{n+1}v = T(T^n y) - T(T^n v) = U(v)(T^n y-T^n v) = U(v)^{n+1}(y-v)$, which proves the equality for n+1.

(b) Since $\|x(\lambda)-\tilde{L}(x)\|_d = \lambda\|x-\tilde{L}(x)\|_d \le \delta(\tilde{L}(x))$ for $\lambda \in [0,1]$, it follows from part (a) with $v = \tilde{L}(x)$ that $T^n x(\lambda) - ng^* - \tilde{L}(x) = U(v)^n(x(\lambda) - \tilde{L}(x)) = U(v)^n(\lambda(x-\tilde{L}(x))) = \lambda U(v)^n(x-\tilde{L}(x))$, the last equality following from (3.5). Since, $U(v)^n(x-\tilde{L}(x)) = T^n x - ng^* - \tilde{L}(x)$, part (b) follows by letting n tend to infinity.

(c) Since for $0 \le \lambda < \delta(v)$, $\|(v+\lambda p)-v\|_d < \delta(v)$, it follows from part (a) and (3.5) that $T^n(v+\lambda p) - (ng^*+v) = \lambda U(v)^n p$. The assertion follows again, by letting n tend to infinity.

4. GEOMETRIC CONVERGENCE IN PHASE 2 AND PHASE 3.

Thanks to lemma 2.2, part (d), the behaviour of $\{v(n)-ng^*\}_{n=1}^{\infty}$ for $v(0) \in W$ in phase 2 and phase 3 can be studied by considering the convergence of $\{T^n x - ng\}_{n=1}^{\infty}$ for $x \in \tilde{W}$. Since for $x \in V$, $x = T^n x - ng^* = \tilde{L}(x)$ for all $n = 1,2,\dots$ we can in general restrict ourselves to (cf. lemma 2.2 part (h)):

$$W^* = \tilde{W}\backslash V = \{x \in \tilde{W} \mid \|\tilde{e}(0,x)\|_d = \|x-\tilde{L}(x)\|_d > 0\}$$

Since $\|\tilde{e}(n,x)\|_d$ is monotonically non-increasing (cf. lemma 2.2 part (e)) we will consider for $n = 1,2,\dots$ the n-step contraction factor $f_n(x)$, defined by:

$$(4.1) \qquad f_n(x) = \begin{cases} \dfrac{\|\tilde{e}(n,x)\|_d}{\|\tilde{e}(0,x)\|_d} = \dfrac{\|T^n x - ng^* - \tilde{L}(x)\|_d}{\|x-\tilde{L}(x)\|_d} = \dfrac{\|T^n x - T^n\tilde{L}(x)\|_d}{\|x-\tilde{L}(x)\|_d}, & \text{for } x \in \tilde{W} \\ 0 & \text{for } x \in V \end{cases}$$

the last equality following from parts (b) and (g) of lemma 2.2. Observe using lemma 2.2 part (e) that $o \le f_{n+1}(x) \le f_n(x) \le 1$ for all $n = 1,2,\ldots$ and that for fixed n, $f_n(x)$ is a continuous function on W^* (cf. lemma 2.1 part (d)). We now prove our main result:

THEOREM 4.1. *There exists an integer* $M \ge 1$ *such that* $f_M(x) < 1$, *for every* $x \in \tilde{W}$.

PROOF. Define:

$$W_A^* = \{x \in W^* \mid \tilde{e}(o,x)_{max} > 0 \text{ and } \tilde{e}(o,x)_{min} \le 0\}$$
$$W_B^* = \{x \in W^* \mid \tilde{e}(o,x)_{max} = 0 \text{ and } \tilde{e}(o,x)_{min} < 0\}$$

Note, using (2.15) that $W^* = W_A^* \cup W_B^*$. Define for $x \in W^*$, $S_n(x) = \{i \mid \tilde{e}(n,x)_i = \tilde{e}(o,x)_{max}\}$. It follows from lemma 2.1 part (g) that:

$$(4.2) \quad S_{n+1}(x) = \{i \mid \text{there exists an alternative } k \in L(i,\tilde{L}(x)), \text{ such that } \sum_{j \in S_n(x)} P_{ij}^k = 1\}$$

For any $v \in V$ define the set of pure policies $SP(v) = X_{i=1}^N L(i,v)$. Note that there exists a finite sequence $\{v^{(1)},\ldots,v^{(R)}\}$ such that $U_{v \in V} SP(v) = U_{\ell=1}^R SP(v^{(\ell)})$.

Let $\{\Omega^{(k)};\ k = 1,\ldots,2^N-1\}$ be the finite collection of non-empty subsets of Ω, and define the following partition of W_B^*.

$$W_{\ell,m}^* = \{x \in W_B^* \mid SP(\tilde{L}(x)) = SP(v^{(\ell)}), S_0(x) = \Omega^{(m)}\}, \ell = 1,\ldots,R;\ m = 1,\ldots,2^N-1.$$

Finally let $I(x) = \inf\{n \mid \|\tilde{e}(n,x)\|_d < \|\tilde{e}(0,x)\|_d\}$, which is finite, for $x \in W^*$, since $\lim_{n\to\infty} \tilde{e}(n,x) = 0$.

In part I) below we show $\sup_{x\in W_A^*} I(x) < 2^N-1$ and in part II) $\sup_{x\in W^*_{\ell,m}} I(x) < \infty$ for fixed $1 \le \ell \le R$ and $1 \le m \le 2^N - 1$, which together imply the theorem:

I) Since $\lim_{n\to\infty} \tilde{e}(n,x)_{max} = 0$, for each $x \in W_A^*$ let $I_0(x)$ be the smallest integer such that $S_n(x)$ is empty for $n \ge I_0(x)$.
Now, $\|\tilde{e}(I_0(x),x)\|_d < \|\tilde{e}(0,x)\|_d$. In addition, in the sequence $\{S_0(x),\dots,S_{I_0(x)-1}\}$ no two members can be equal since using (4.2) this would imply that $S_n(x)$ is non-empty for all $n \ge 1$.
Hence $I_0(x) \le 2^N - 1$ since there are only $2^N - 1$ distinct non-empty subsets of Ω.

II) Fix $x^0 \in W^*_{\ell,m}$. Due to (3.1) and (3.2) there exists an integer N_1 such that $\tilde{e}(n+1,x^0)_i = [P(f_n)\dots P(f_{n_1+1})\tilde{e}(n_1,x^0)]_i$ for $i = 1,\dots,N$; $n \ge n_1+1$ where $f_n,\dots,f_{n_1+1} \in SP(v^{(\ell)})$. Define I_1 as follows:

$$I_1 = \begin{cases} \min\{n\ge n_1+1 \mid 0 \ge \tilde{e}(n,x^0)_{min} > \tilde{e}(n_1,x^0)^-\} & \text{if } S_{n_1}(x^0) \ne \Omega. \\ n_1+1 & \text{otherwise} \end{cases}$$

Then in both cases I_1 is finite, since n_1 is finite and $\lim_{n\to\infty}\tilde{e}(n,x^0)_{min} = 0$. In addition, we shall prove for both cases:

$$\sum_{j\in S_{n_1}(x^0)} [P(f_{I_1})\dots P(n_1)]_{ij} > 0, \text{ for all } i \in \Omega. \tag{4.3}$$

(4.3) trivially holds if $S_{n_1}(x^0) = \Omega$, and for the other case we have $\tilde{e}(I_1,x^0)_{min} \le \tilde{e}(n_1,x^0)^-$ if (4.3) does not hold. This contradicts the definition of I_1. Next fix for $r = 1,\dots,n_1+1$, $f_r \in SP(v^{(\ell)})$ such that

$$\sum_{j\in S_0(x^0)=\Omega(m)} [P(f_{I_1})\dots P(f_1)]_{ij} > 0, \text{ for all } i \in \Omega, \tag{4.4}$$

the existence of which follows from $S_{n_1}(x^0) \ne \emptyset$ in view of $\tilde{e}(0,x)_{max} = 0$ in combination with lemma 2.1 part (g).
Now observe that for all $x \in W^*_{\ell,m}$ we have $b(\tilde{L}(x),f_n) = 0$ for

$n = 1,\dots,I_1$ since $f_n \in SP(v^{(\ell)}) = SP(\tilde{L}(x))$. Hence, using (3.1), (3.2) and (4.4) $\tilde{e}(I_1,x) \geq \sum_{j\in\Omega-S_0(x)} [P(f_{I_1})\dots P(f_1)]_{ij}\ \tilde{e}(0,x)_j > \tilde{e}(0,x)_{min}$. This implies that for all $x \in W^*_{\ell,m} : I(x) < I_1$. □

In order to prove the geometric convergence of $\{T^n x-ng^*\}_{n=1}^{\infty}$, we define:

(4.5) $$h_m(x) = \sup_{n=0,1,\dots} f_m(T^n x), \quad x \in \tilde{W} \text{ and } m = 0,1,\dots$$

which has the following easily verified properties:

(4.6) $$h_m(x) = h_m(x+c_1 g^* + c_2 \underline{1}), \text{ for all scalars } c_1,c_2;\ x \in \tilde{W};\ m = 0,1,\dots$$
$$0 \leq h_{m+1}(x) \leq h_m(x) \leq 1, \quad x \in \tilde{W};\ m = 0,1,\dots$$
$$h_m(T^r x) \leq h_m(x), \quad x \in \tilde{W};\ m,r = 0,1,\dots$$

THEOREM 4.2. (Geometric convergence result).

(a) $h_m(x) < 1$ *for all* $m \geq M$ *and* $x \in \tilde{W}$.

(b) $\|\tilde{e}(nM+r,x)\|_\infty \leq \|\tilde{e}(nM+r,x)\|_d \leq [h_M(x)]^n \|\tilde{e}(0,x)\|_d$ for $n = 0,1,2,\dots$; $r = 0,1,\dots;M-1$ *and* $x \in \tilde{W}$.

Hence the convergence of $\{T^n x-ng^*\}_{n=1}^{\infty}$ *is geometric for all* $x \in \tilde{W}$.

PROOF.

(a) Suppose to the contrary that $h_M(x) = 1$ for some $x \in \tilde{W}$. It then follows from (4.1) and lemma 2.2 part (b), that $x \in W^*$ and that there exists a subsequence $\{x^j\}_{j=1}^{\infty} = \{T^{n_j}x - n_j g^*\}_{j=1}^{\infty}$ such that $\lim_{j\to\infty} f_M(x^j) = 1$. Using lemma 2.1 part (f), it easily follows that $x^j \in \tilde{W}$ with $\tilde{L}(x^j) = \tilde{L}(x)$ and $\|x^j-\tilde{L}(x)\|_d > 0$ for all $j = 1,2,\dots$
Put $x^j = \tilde{L}(x) + \xi^j$. Since for j large enough, $\|x^j-\tilde{L}(x)\|_d < \delta(\tilde{L}(x))$, we have using lemma 3.2 part (a), for all $n \geq 1$:
$T^n(x^j) = \tilde{L}(x) + ng^* + U(\tilde{L}(x))^n(\xi^j)$, and $\lim_{n\to\infty} U(\tilde{L}(x))^n(\xi^j) = 0$ for j sufficiently large. Hence,

$$1 = \lim_{j\to\infty} f_M(x^j) = \lim_{j\to\infty} \frac{\| T^M(x^j) - Mg^* - \tilde{L}(x)\|_d}{\| x^j - \tilde{L}(x)\|_d} =$$

$$\lim_{j\to\infty} \frac{\| U(\tilde{L}(x))^M(\xi^j)\|_d}{\|\xi^j\|_d}$$

For any $v \in V$, define $Y(v)=\{y \in E^N \mid \|y\|_d = 1 \text{ and } \lim_{n\to\infty} U(v)^n y=0\}$, and

(4.7) $$\Gamma_n(v) = \begin{cases} \sup_{y\in Y(v)} \|U(v)^n y\|_d & \text{if } Y(v) \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

Observing with the help of (3.3) that $\xi^j/\|\xi^j\|_d \in Y(\tilde{L}(x))$, $j = 1,2,\ldots$ and recalling that $\Gamma_n(v) \le 1$, $n = 1,2,\ldots$ and $v \in V$ (cf. lemma 2.1 part (b)), we conclude that $\Gamma_M(\tilde{L}(x)) = 1$.

Observe by lemma 2.1 part (e) that $Y(v)$ is closed for any $v \in V$. In addition $Y(v)$ is bounded since for any $y \in Y(v)$, $y_{max} \ge 0 \ge y_{min}$ as a result of lemma 2.2 part (e) being applied to the $U(v)$-operator, and hence $\|y\|_\infty \le \|y\|_d = 1$ for any $y \in Y(v)$ (cf.(2.9)).
We conclude that in (4.7) the supremum is taken of a continuous function (cf. lemma 2.1 part (d)) over a compact set, and this implies the existence of a vector $y^0 \in Y(\tilde{L}(x))$ with $\|U(\tilde{L}(x))^M y^0\|_d = 1$.
Invoking lemma 3.2 part (c) we find that $\tilde{L}(x) + \lambda y^0 \in W^*_,$ for $0 < \lambda < \delta(\tilde{L}(x))$ with $\tilde{L}(\tilde{L}(x) + \lambda y^0) = \tilde{L}(x)$. Next using lemma 3.2 part (a) and (3.3):

$$f_M(\tilde{L}(x)+\lambda y^0) = \frac{1}{\lambda} \|T^M(\tilde{L}(x)+\lambda y^0) - T^M(\tilde{L}(x))\|_d = \frac{1}{\lambda} \|U(\tilde{L}(x))^M(\lambda y^0)\|_d = 1,$$

thus contradicting th. 4.1.

(b) Fix $x \in \tilde{W}$, $n = 0,1,\ldots$ and $1 \le r \le M$:
The first inequality follows from part (c) of lemma 2.2 and (2.9).
If $\|\tilde{e}(nM+r,x)\|_d = 0$, we trivially have:

(4.8) $\|\tilde{e}((n+1)M+r,x)\|_d \le h_M(x)\ \|\tilde{e}(nM+r,x)\|_d$

Next assume $\|\tilde{e}(nM+r,x)\|_d > 0$. Then

$$\frac{\|\tilde{e}(nM+M+r,x)\|_d}{\|\tilde{e}(nM+r,x)\|_d} = f_M(T^{nM}x) \le h_M(T^{nM}x) \le h_M(x),$$

the last inequality following from (4.6).
This proves the second inequality in part (b) for all $x \in \tilde{W}, n = 0,1,\ldots$ and $r = 1,\ldots,M$. □

Th.4.2 in combination with lemma 2.2 part (d) establish the geometric convergence result for all $x \in W$. If $x \notin W$, then certain subsequences of the type:

(4.9) $\{Q^{nJ+r}x - (nJ+r)g^*\}_{n=1}^{\infty}; J = 2,3,\ldots$ and $r = 0,\ldots,J-1$

will converge. We refer to th. 5.8 of [18] for a characterization of the integers $J \ge 1$ for which convergence occurs. Fix $J = 2,3,\ldots$ and note that: (cf. section 4 in [18]):

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

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

$$\tilde{q}_i^\xi = q(f^1)_i + P(f^1)q(f^2)_i + \ldots + P(f^1)\ldots P(f^{J-1})q(f^J)_i, \quad i \in \Omega,\ \xi = (f^1,\ldots,f^J) \in \tilde{K}(i)$$

$$\tilde{P}_{ij}^\xi = P(f^1)\ \ldots\ P(f^J)_{ij}; \qquad 1 \le i,j \le N \text{ and } \xi = (f^1,\ldots,f^J) \in \tilde{K}(i).$$

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 expexted 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.

Recalling from th. 4.1 part (a) in [18] that $\tilde{g}^* = Jg^*$ we obtain in view of $\{Q^{nJ+r}x-(nJ+r)g^*\}_{n=1}^{\infty} = \{\tilde{Q}^n[Q^r x]-n\tilde{g}^*\}_{n=1}^{\infty} - rg^*$ and by applying the above analysis to the J-step MDP, the following generalization of the geometric convergence result.

COROLLARY 4.3. *Fix* $J = 1,2,\ldots$ *and* $r = 0,\ldots,J-1$.
If $\lim_{n\to\infty} Q^{nJ+r}x - (nJ+r)g^*$ *exists, then the approach to the limit exhibits a geometric rate of convergence.* □

REMARK 2.: Assume $g^* = \langle g^*\rangle \underline{1}$ so $Q = T$ and consider White's iterative scheme for solving MDP's (cf. [22]). Define:

$$y(n)_i = v(n)_i - v(n)_N,\quad i = 1,\ldots,N;$$

and verify that

$$y(n+1) = Qy(n) - [Qy(n)_N]\underline{1}$$

Then if $v(0) \in W = \tilde{W}$:

(a) $\lim_{n\to\infty} y(n)_i = L(v(0))_i - L(v(0))_N$
(b) $[Qy(n) - y(n)]_{max} = [v(n+1) - v(n)]_{max} \downarrow g^*$ $n \to \infty$ (cf. ODONI [12], th.1).
(c) $[Qy(n) - y(n)]_{min} = [v(n+1) - v(n)]_{min} \uparrow g^*$ $n \to \infty$ (cf. ODONI 12 , th.1).

It follows from th. 4.2 that the convergence in (a), (b) and (c) is geometric since $|\, y(nM+r)_i - L(v(0))_i - L(v(0))_N| \le \| e(nM+r,v(0))\|_d \le [h_M v(0)]^n \| e(0,v(0))\|_d$.

5. THE SIZE OF M

In this section we restrict ourselves to MDP's that satisfy the condition:

(H1): there exists a $f^o \in S_{RMG}$ that is aperiodic and has R^* as its single subchain.

In [17] we proved that (H1) is satisfied e.g. if all the tpm's of the pure maximal gain policies are unichained, whereas the greatest common divisor of their periods equals 1.

Fix $v \in V$; we first observe that the policy f^*, defined by:

$$(5.1)\qquad \{k \mid f^*_{ik} > 0\} = \{k \in L(i) \mid b(v)^k_i = 0\},\quad i \in \Omega$$

is one of the policies with the properties mentioned in (H1).
Using (2.8) one first observes that $f^* \in S_{RMG}$ hence $R(f^*) \subseteq R^*$.
Due to (H1) all states of R^* communicate with each other under $P(f^o)$ and since for all $i \in R^*, f^o_{ik} > 0$ implies by (2.8) $k \in L(i)$ and $b(v)^k_i = 0$, hence $f^*_{ik} > 0$ they communicate with each other under $P(f^*)$. Hence $P(f^*)$ is aperiodic and has R^* as its single subchain.

Lemma 5.1 below gives some implications with respect to the chain- and periodicity structure that result from (H1).

LEMMA 5.1. *Suppose* C1 *holds. Then:*

(a) $g^* = \langle g^* \rangle \underline{1}$, i.e. $K(i) = L(i)$ *for all* $i \in \Omega$, *and* $Qx = Tx$ *for all* $x \in E^N$.

(b) $v \in V$ *is unique up to a multiple of* $\underline{1}$.

(c) *For all* $i \in \Omega$, *and* $k \in K(i)$, $b(v)^k_i$ *is independent of* $v \in V$.

(d) $W = \tilde{W} = E^N$.

(e) *If* $v \in V$, $i \in R^*$ *and* $b(v)^k_i = 0$ *then* $P^k_{ij} > 0$ *only if* $j \in R^*$.

(f) *For any bounded subset* $B \subset E^N$: $\sup_{x \in B} f_M(x) < 1$ (*where* M *is defined as in* th.4.1).

PROOF. Parts (a) and (b) follow from th. 3.2 parts (c) and (e) and remark 2 in [17]. Part (c) follows from (2.6) and part (b); part (d) is proven in [18]. To show part (e), suppose there exists (i,j,k) with $i \in R^*$, $j \notin R^*$, $b(v)^k_i = 0$ for $v \in V$ and $P^k_{ij} > 0$. Then $f^*_{ik} > 0$ and $P(f^*)_{ij} \geq f^*_{ik} P^k_{ij} > 0$ contradicting the fact that $R(f^*) = R^*$.

(f): Assume to the contrary that for some bounded subset $B \subset E^N$, $\sup_{x \in B} f_M(x) = 1$. Considering the definition of $f_n(x)\,(n \geq 1)$ we assume without loss of generality that $B \subset W^*$. Then there exists a sequence $\{x^j\}_{j=1}^{\infty}$, with $x^j \in B$ such that $\lim_{j\to\infty} x^j = c \in W$ (say) and $\lim_{j\to\infty} f_M(x^j) = 1$.

The case $c \in W^*$ leads to the contradiction $1 = \lim_{j\to\infty} f_M(x^j) = f_M(c) < 1$ in view of th. 4.1 and the continuity of $f_M(\cdot)$ on W^*.
The remaining case has $c \in V$. Put $x^j = L(x^j) + \xi^j$. Following the proof of th. 4.2 part (a) we obtain for j sufficiently large:

$$T^n(x^j) = v + ng^* + U(v)^n \xi^j \text{ and so } \lim_{n\to\infty} U(v)^n[\xi^j] = L(x^j) - v$$

Since it follows from part (b) that $L(x^j) - v$ is a multiple of $\underline{1}$ we obtain:

$$f_M(x^j) = \frac{\| T^M(x^j) - Mg^* - L(x^j) \|_d}{\| x^j - L(x^j) \|_d} = \| U(v)^M (y^j) \|_d$$

where

$y^j = (\xi^j + v - L(x^j)) / \| \xi^j \|_d \in Y(v)$. The remainder of the proof is completely analoguous to that of th. 4.2 part (a). □

We next derive (for MDP's satisfying (H1)) an upperbound for M the number of steps needed for contraction:
First define:

(5.1) $$\gamma = \min\{ n \geq N \mid P(f^*)^n_{ij} > 0, \quad \text{for all } i = 1,\ldots,N, \quad j \in R^* \}$$

Clearly $\gamma < \infty$, since $\lim_{n\to\infty} P(f^*)^n_{ij} > 0$ for all $i = 1,\ldots,N$ and $j \in R^*$. Note that $P(f^*)^n_{ij} > 0$ for all $i \in \Omega$, $j \in R^*$ and $m \geq \gamma$, since for $m \geq \gamma$ $P(f^*)^m_{ij} = \sum_{k=1}^N P(f^*)^{m-\gamma}_{ik} \cdot P(f^*)^\gamma_{kj} > 0$ for all $i \in \Omega, j \in R^*$.

<u>THEOREM 5.2</u>. *If* (H1) *holds then* $M \leq N^2 - 2N + 2$, (*where* M *is defined as the smallest integer satisfying the condition of* th.4.1.).

<u>PROOF</u>. We will first show that $\gamma \leq N^2 - 2N + 2$. Assume that $R^*_i R(f^*)$ contains $N + k \geq 1$ states. Then it follows from th. 2.8 of [17] that $P(f^*)^n_{ij} > 0$ for $n \geq (N-k)^2 - 2(N-k) + 2$ and $i,j \in R^*$. In addition for any $i \in \Omega - R^*$, there exists a path $\{t_0 = i, t_1, \ldots, t_m\}$ such that $P(f^*)t_\ell t_{\ell+1} > 0$ for $\ell = 0,\ldots,$ $m-1$ and $t_m \in R^*$, where without loss of generality $t_1,\ldots,t_m$ are all taken to be distinct. Hence $m \leq k$ and $\sum_{\ell\in R^*} P(f^*)^k_{i\ell} > 0$ for all $i \in \Omega$.
This implies that $P(f^*)^n_{ij} \geq \sum_{\ell\in R^*} P(f^*)^k_{i\ell} P(f^*)^{n-k}_{\ell j} > 0$ for all $i \in \Omega$, $j \in R^*$ and $n \geq N^2 - 2N + 2$ (verify that $k + (N-k)^2 - 2(N-k) + 2 \leq N^2 - 2N + 2$ in view of the quadratic form $(3-2N)k + k^2$ being nonpositive for $k = 0,\ldots,N-1$).

Next we fix $x \in W^*$. Let $L(x) = v^*$ and define:

$$X(m) = \{ i\in\Omega \mid (T^m x - T^m v^*)_i = (x - v^*)_{max} \};\ m=0,1,2,\ldots$$
$$Y(m) = \{ i\in\Omega \mid (T^m x - T^m v^*)_i = (x - v^*)_{min} \};\ m=0,1,2,\ldots$$

We will prove that $M \leq \gamma$ and hence $M \leq N^2 - 2N + 2$, by showing that the assumption $M > \gamma$ implies (a) $Y(0) \supseteq R^*$ and (b) $X(0) \cap R^* \neq \emptyset$, hence $X(0) \cap Y(0) \neq \emptyset$ contradicting $x \in W^*$, i.e. $\|x-v^*\|_d > 0$. Assume now $\gamma < M$. Then $X(m) \neq \emptyset \neq Y(m)$ for $0 \leq m \leq \gamma$. Fix $m \leq \gamma$, and $i \in Y(m)$. Observe using part (h) of lemma 2.1, that for any $k \in L(i,v^*)$ $P^k_{ij} > 0$ only if $j \in Y(m-1)$. Using the definition of f^*, we conclude that $P(f^*)_{ij} > 0$ only if $j \in Y(m-1)$. Proceeding by induction, and invoking the definition of γ we obtain for $i \in Y(\gamma)$: $R^* \subseteq \{j \mid P(f^*)_{ij} > 0\} \subseteq Y(0)$.

The nested sequence $X(N)$; $X(N) \cup X(N-1)$;...; $\cup_{i=0}^{N} X(i)$ cannot exhibit strict growth since there are only N states, hence there exists a $m \leq N - 1$ such that $X(m) \subset S = \cup_{\ell=m+1}^{N} X(\ell)$. Accordingly define a policy h in the following way:

(a) for $i \in \Omega - S$, define $h(i) = k$ for some $k \in L(i,v^*)$

(b) for $i \in S$, choose an index $\ell (m+1 \leq \ell \leq N)$ such that $i \in X(\ell)$, and define $h(i) = k$ for any $k \in L(i,v^*)$ such that $P^k_{ij} > 0$ only if $j \in X(\ell-1)$, the existence of such an alternative k being guaranteed by part (g) of lemma 2.1, and the fact that $L(i,T^{\ell}v^*) = L(i,v^*)$.

It clearly follows from (2.6) that $h \in S_{PMG}$; in addition S contains a subchain of P(h) since it follows from $X(m) \subset S$, that S is closed under P(h). Hence, $S \cap R^* \neq \emptyset$, or there exists an index r, such that $X(r) \cap R^* \neq \emptyset$ Accordingly fix $i \in X(r) \cap R^*$. Then, again applying part (g) of lemma 2.1 we obtain the existence of an alternative $k \in L(i,v^*)$ such that $P^k_{ij} > 0$ only for $j \in X(r-1)$.

In addition, since $i \in R^*$ and $k \in L(i,v^*)$ it follows from lemma 5.1 part (e) that $P^k_{ij} > 0$ only for $j \in R^*$. Hence $X(r) \cap R^* \neq \emptyset$ implies $X(r-1) \cap R^* \neq \emptyset$ and proceeding by induction we obtain $X(0) \cap R^* \neq \emptyset$. This together with $Y(0) \supseteq R^*$ implies $X(0) \cap Y(0) \neq \emptyset$, i.e. $\|x-L(x)\|_d = 0$ thus contradicting $x \in W^*$. □

The following example shows that $M = O(N^2)$ may occur.

Example 2:

i	k	q_i^k	P_{i1}^k	P_{i2}^k	P_{i3}^k	...	P_{iN-1}^k	P_{iN}^k
1	1	0	0	1				
2	1	0	0	0	1			
N-2	1	0					1	
N-2	2	0						1
N-1	1	0	$\frac{1}{2}$	$\frac{1}{2}$				
N	1	0	1					

$K(i) = \{1\}$ for $i \neq N-2$; $K(N-2)=\{1,2\}$; $P_{ii+1}^1=1$ for $i \leq N-2$; $q_i^k = 0$ for all i,k; hence $g^* = 0$ and $K(i) = L(i)$ for all $i \in \Omega$.

Let $f_k (k=1,2)$ denote the pure policy that chooses alternative k in state N-2. Observe that (H1) holds since $P(f_1)$ and $P(f_2)$ are unichained with $P(f_1)$ aperiodic. Consider x, with $x_i = 0$ for $i \neq N-1$ and $x_{N-1} = 1$.

Clearly $[T^{J(N-1)}x]_N = [P(f_2)^{(J-1)(N-1)}P(f_1)^{N-1}x]_N = 1$, for $J = 1,2,\ldots$. Observe that whatever decisions are taken when entering state N-2, the only states j that can be reached from state 1, after J(N-1) steps are $j = 1,\ldots,J+1$ $(J \leq N-1)$. Hence $[T^{(N-3)(N-1)}x]_1 = 0$.

Note, using lemma 5.1, parts (b) and (c) that $x \in W^*$ with $L(x) = \lambda \underline{1}$ for some scalar λ. Hence, $\|T^{(N-1)(N-3)}x - L(x)\|_d = [T^{(N-1)(N-3)}x]_N - [T^{(N-1)(N-3)}x]_1 = 1 = \|x-L(x)\|_d$, and $M \geq (N-3)(N-1)$.

<u>REMARK 3</u>. The upperbound $N^2 - 2N + 2$ for the number of iterations needed for contraction is enormously high, compared with the empirical fact that in most cases M = 1 or 2. For example SU [20] and TIJMS [21] have solved up to 1000-state problems with good convergence after 10 - 100 value iterations. In addition if $P(f^*)$ has at least one positive diagonal entry, it may be shown that the upperbound for M becomes *linear* in N.
Since it was shown in [8] that in this case $\gamma \leq 2N - r - 1$, where $r \geq 1$ is the number of positive diagonal entries of $P(f^*)$ the result $M = O(2N)$ again follows from the proof of th.5.2.
In SCHWEITZER [6] a data-transformation was introduced which turns every MDP into an equivalent one in which all of the diagonal elements of the tpm's are positive thus ensuring convergence of $\{Q^n x - ng^*\}_{n=1}^{\infty}$, for all $x \in E^N$. By the above analysis it follows that thanks to this transforma-

tion, M the number of steps needed for contraction, is in addition bounded by N-1. Finally in case S_P consists of a single unichained and aperiodic policy, we have $M \le \frac{1}{2}N(N-1)$ as a result of the following argument:
We know (cf. th.4.4 on pp. 89 of [19]) that any aperiodic and unichained policy f, has $P(f)^n$ *scrambling* for all $n \ge \frac{1}{2}N(N-1)$, i.e.
$\min_{i_1,i_2} \sum_j \min[P(f)^n_{i_1 j};P(f)^n_{i_2 j}] = \alpha > 0$ for all $n \ge \frac{1}{2}N(N-1)$.

One next verifies (cf. th.5 in [7]) that $\|e(n,x)\|_d \le (1-\alpha)\|e(0,x)\|_d$ for all $x \in E^N$ and $n \ge \frac{1}{2}N(N-1)$.

6. THE THIRD PHASE; THE ULTIMATE CONVERGENCE RATE

In this section we analyze the ultimate convergence rate or average contraction factor per step which is defined as the limit as n tends to infinity of:

$$(6.1)\quad f_n(x)^{1/n} = \begin{cases} \left[\dfrac{\|\tilde{e}(n,x)\|_d}{\|\tilde{e}(n-1,x)\|_d} \dfrac{\|\tilde{e}(n-1,x)\|_d}{\|\tilde{e}(n-2,x)\|_d} \cdot \ldots \cdot \dfrac{\|\tilde{e}(1,x)\|_d}{\|\tilde{e}(o,x)\|_d}\right]^{1/n}, & \text{if } \|\tilde{e}(n-1,x)\|_d > 0 \\ 0 & \text{otherwise} \end{cases}$$

Note that $f_n(x)^{\frac{1}{n}}$ may be interpreted as the (geometric) mean n-step contraction factor. In section 3 we observed that for $n \ge n_1(x)$ (cf. (3.1)) i.e. in the third phase, the sequence $\{e(n,x)\}_{n=1}^{\infty}$ satisfies the recursion equation:

$$(6.2)\qquad e(n+1,x) = U(L(x))e(n,x), \quad x \in W; \quad n \ge n_1(x)$$

Thus, in order to characterize the ultimate convergence rate, the following two theorems give some properties of the U-operator and of the quantities $\Gamma_n(v); v \in V$:

$$(6.1)\qquad \Gamma_n(v) = \begin{cases} \sup_{y \in Y(v)} \|U(v)^n y\|_d & \text{if } Y(v) \ne \emptyset \\ 0 \end{cases}$$

where

$$Y(v) = \left\{ y \in E^N \mid \|y\|_d = 1; \lim_{n\to\infty} U(v)^n y = 0 \right\}$$

First, define for all $v \in V$, $W_{U(v)} = \{y \in E^N \mid \lim_{n\to\infty} U(v)^n y \text{ exists}\}$, and for all $y \in W_{U(v)}$, let $U(v)^\infty y = \lim_{n\to\infty} U(v)^n y$.

THEOREM 6.1.

(a) (Cf.th.4.1). *There exists an integer* $M_1 \le 2^N$ *such that for all* $v \in V$ *and* $y \in W_{U(v)}$ *with* $\|y-U(v)^\infty y\|_d > 0$:

$$\|U(v)^{M_1} y - U(v)^\infty y\|_d < \|y-U(v)^\infty y\|_d$$

Fix $v \in V$.

(b) *If* $Y(v) \neq \emptyset$ *then* $\Gamma_n(v) = \max_{y\in Y(v)} \|U(v)^n y\|_d$; $n = 1,2,\ldots$

(c) $\|U(v)^{M_1} y\|_d \le (1-\rho_1)\|y\|_d$, *for all* $y \in E^N$ *such that* $U(v)^\infty y = 0$, *where*

$$1-\rho_1 = \max_{v\in V} \Gamma_{M_1}(v) < 1 \tag{6.2}$$

(d) $\Gamma_{m+n}(v) \le \Gamma_m(v).\Gamma_n(v)$ *for all* $m,n = 0,1,2,\ldots$

(e) *Define* $\Gamma^*(v) = \lim_{n\to\infty} \Gamma_n(v)^{1/n}$. *Then* $\Gamma^*(v) \le (1-\rho_1)^{1/M_1} < 1$ *and* $\Gamma_n(v) \ge \Gamma^*(v)$ *for all* $n = 0,1,\ldots$.

PROOF.

(a) Fix $v \in V$ and $y \in W_{U(v)}$ and recall from lemma 2.2 part (e) that $(y-U(v)^\infty y)_{min} \le 0 \le (y-U(v)^\infty y)_{max}$. Define for $n = 1,2,\ldots$:

$$S_n = \{i \mid U(v)^n(y-U(v)^\infty y)_i = (y-U(v)^\infty y)_{max}\}$$

and

$$T_n = \{i \mid U(v)^n(y-U(v)^\infty y)_i = (y-U(v)^\infty y)_{min}\}.$$

Observe using the arguments in part I) of the proof of th.4.1 that S_n must be empty for $n \ge 2^N$ if $(y-U(v)^\infty y)_{max} > 0$. However, for the U-operator the same arguments show that T_n must be empty for $n \ge 2^N$ if $(y-U(v)^\infty y)_{min} < 0$, as well.

(b) In the proof of th.4.1. part (b) we showed that the supremum in (4.6) is always achieved by some $y^o \in Y(v)$.

(c) It follows from part (a) and (b) that $\Gamma_{M_1}(v) < 1$ for any $v \in V$. Since there are only a finite number of distinct U(v)-operators, we have $\max_{v\in V}\Gamma_{M_1}(v) < 1$, which proves (6.2) and hence the remainder of part (e).

(d) For $y \in Y(v)$ with $\|U(v)^n y\|_d=0$, we have:

$$0 = \|U(v)^{n+m}y\|_d \le \Gamma_m(v)\Gamma_n(v)$$

while for $y \in Y(v)$, with $\|U(v)^n y\|_d > 0$:

$$\|U(v)^{n+m}y\|_d = \left\| U(v)^m \left\{\frac{U(v)^n y}{\|U(v)^n y\|_d}\right\} \right\|_d \|U(v)^n y\|_d \le \Gamma_m(v)\Gamma_n(v)$$

Hence $\Gamma_{n+m}(v) = \max_{y\in Y(v)} \|U(v)^{n+m}y\|_d \le \Gamma_m(v)\Gamma_n(v)$.

(e) The existence of $\Gamma^*(v) = \lim_{n\to\infty}\Gamma_n(v)^{1/n}$ and the relation $\Gamma^*(v) \le \Gamma_n(v)^{1/n}$ for all $n = 1,2,\ldots$ follows from part (d) and a well-known theorem of KINGMAN (cf. e.g. [19], appendix A, th. A4).
It follows from (6.2) that $\Gamma_{M_1}(v) \le 1-\rho_1$, and hence using part (d), that $\Gamma_{nM_1}(v) \le (1-\rho_1)^n$. This implies:

$$\Gamma^*(v) = \lim_{n\to\infty} \Gamma_{nM_1}(v)^{1/nM_1} \le (1-\rho_1)^{1/M_1}. \quad \square$$

Th. 6.2. below proves that for any $x \in W^*$ the ultimate average contraction factor per step is at worst $\Gamma^*(L(x))$, so that for all $x \in W^*$, the ultimate convergence rate is strictly bounded away from one.
In addition, part (b) shows that for any *fixed* n, there are $x \in W^*$ for which the average n-*step* contraction factor is at least equal to $\max_{v\in V}\Gamma_n(v)^{1/n}$.

THEOREM 6.2.

(a) $\limsup_{n\to\infty} f_n(x)^{1/n} \le \Gamma^*(\tilde{L}(x))$ *for any* $x \in W^*$.

(b) $\sup_{x\in\tilde{W}} f_n(x)^{1/n} \ge \max_{v\in V} \Gamma_n(v)^{1/n} \ge \max_{v\in V} \Gamma^*(v)$, *for all* $n = 0,1,\ldots$.

PROOF.

(a) Fix $x \in \tilde{W}$ and observe that by (4.1):
$f_{n+m}(x) = f_m(T^n x-ng^*)\, f_n(x)$. Fix n sufficiently large that $\|T^n x-ng^*-\tilde{L}(x)\|_d < \delta(\tilde{L}(x))$.

Then, either $T^n - ng^* = \tilde{L}(x)$ in which case $f_m(x) = 0$ for all $m \geq n$ and part (a) trivially holds, or otherwise we have, using lemma 3.1 part (a) and (3.4):

$f_m(T^n x - ng^*) = U(\tilde{L}(x))^m y$, where $y = (T^n x - ng^* - \tilde{L}(x))/\| T^n x - ng^* - \tilde{L}(x)\|_d$.

Hence, in the latter case $f_{n+m}(x) \leq \Gamma_m(\tilde{L}(x)) f_n(x)$, or

$\limsup_{m\to\infty} f_{n+m}(x)^{1/n+m} \leq \lim_{m\to\infty} \Gamma_m(\tilde{L}(x))^{1/m+n} \lim_{m\to\infty} f_n(x)^{1/n+m} = \Gamma^*(\tilde{L}(x))$.

(b) Fix $v \in V$. If $Y(v)$ is empty then $\sup_{x\in\tilde{W}} f_n(x)^{1/n} \geq \Gamma_n(v)^{1/n} = \Gamma^*(v) = 0$ holds trivially.

Otherwise considering th. 6.1 part (b), take $y \in Y(v)$ such that $\Gamma_n(v) = \| U(v)^n y\|_d$. Let $x^0 = v + \lambda y$ with $0 < \lambda < \delta(v)$. Then using lemma 3.2 parts (a) and (c) as well as (3.3), we have $x^0 \in \tilde{W}$, $\tilde{L}(x^0) = v$ and:

$f_n(x^0) = \| U(v)^n(\lambda y)\|_d / \| \lambda y\|_d = \| U(v)^n y\|_d / \| y\|_d = \Gamma_n(v)$, or $f_n(x^0)^{1/n} = \Gamma_n(v)^{1/n}$ from which the first inequality of part (b) follows

The second inequality is due to th. 6.1 part (e). □

We conclude this section by observing that the upperbound

(6.3) $$\max_{v\in V} \Gamma_{M_1}(v)^{1/M_1} = \max_{v\in V} \max\{\| U(v)^{M_1} y\|_d^{1/M_1} \mid \| y\|_d = 1, U(v)^\infty y = 0\}$$

for the ultimate convergence rate reduces in the special case where S_{PMG} is a singleton, to the subdominant eigenvalue of the tpm of the maximal gain policy; and in this case the subdominant eigenvalue is known to provide a sharp upperbound for the convergence rate (cf. e.g. [11]).

7. THE N-STEP CONTRACTION FACTOR

Theorem 6.2 showed that $\max_{v\in V}\Gamma^*(v)$ is at the same time an *upperbound* for the *ultimate convergence rate* and a *lower bound* for the maximal average *n-step contraction factor* for all integers n = 1,2,... .

The following example shows that whereas the ultimate convergence rate is strictly bounded away from one, this does not need to be the case for the average n-step contraction factor (whatever the choice of n = 1,2,...).

In other words we may have, *for all* n = 1,2,...:

$$\sup_{x \in W} f_n(x) = 1.$$

EXAMPLE 3:

i	k	q_i^k	P_{i1}^k	P_{i2}^k
1	1	0	1	0
2	1	0	1	0
2	2	-1	0	1

$g^* = (0,0)$ hence $K(i) = L(i)$ for $i = 1,2$
$V = \{\lambda \underline{1} \mid \lambda \text{ arbitrary}\}$. Take $x = [0,Y]$.

Observe that this MDP satisfies condition (H1) (cf. section 5); hence, using lemma 5.1 part (d), we have $\tilde{W} = E^N$:

$$T^n x = [0, \max(0, Y-n)]$$

$$f_n(x) = \|T^n x - ng^* - \tilde{l}(x)\|_d / \|x - \tilde{l}(x)\|_d = \|T^n x\|_d / \|x\|_d =$$

$$\frac{\max\ (0, Y-n)}{Y}$$

Letting Y tend to infinity one observes that $\sup_{x \in \tilde{W}} f_n(x) = 1$ for all $n = 1,2,\ldots$.

The following theorem gives under condition (H1) the necessary and sufficient condition for the existence of a uniform (n-step) contraction factor (for some $n \geq 1$) i.e. the existence of an integer M_2, such that

(7.1) $$\sup_{x \in W} f_n(x) < 1 \text{ for } n \geq M_2.$$

First define:

(7.2) $$\hat{R} = \{i \in \Omega \mid i \in R(f), \quad \text{for some } f \in S_P\}$$

and note that $\hat{R} \supseteq R^*$. We next introduce the condition:

(H2): There exists a randomized policy $f \in S_R$ which has $\hat{R}$ as its single subchain.

THEOREM 7.1. *Suppose condition* (H1) *holds.*

(a) *The existence of a uniform* n*-step contraction factor some* $n \geq 1$ *implies* (H2).

(b) (H2) $\Rightarrow$ (7.1) *with* $M_2 \leq N^2 + 2$.

PROOF: Fix $v \in V$. Due to lemma 5.1, parts (b) and (d) we have $W = E^N$ and for all $x \in E^N$, $L(x) = v + c\underline{1}$ for some scalar c. This implies $W^* = \{x \mid \|x-v\|_d > 0\}$

(a) Assume to the contrary that (H2) does not hold. State i is said to *reach* state j, if there exists a policy $f \in S_P$, and some integer $r \geq 0$, such that $P(f)^r_{ij} > 0$. Let f^* be any randomized policy which has $f^*_{ik} > 0$ for all $i \in \Omega, k \in K(i)$. We claim

(7.3) there exists a pair of states $j_1, j_2 \in \hat{R}$ such that j_2 does not reach j_1.

For assuming the contrary, would imply that all states in $\hat{R}$ communicate with each other under $P(f^*)$, i.e. either

(1) $\hat{R} \subseteq \Omega \backslash R(f^*)$, or

(2) $\hat{R}$ is a strict subset of $R(f^*)$, or

(3) $P(f^*)$ has $\hat{R}$ as a single subchain,

with each of these three possiblities leading to a contradiction in view of the definition of $\hat{R}$, and our assumption that (H2) does not hold. Fix a policy $f_1 \in S_P$ with $j_1 \in R(f_1)$ and let C be the subchain of $P(f_1)$, which contains j_1. Obviously j_2 does not reach any one of the states in C. Next choose $x \in E^N$ such that $x_i = \lambda >> 1$ for $i \in C$ and $x_i = O(1)$ otherwise where $O(1)$ denotes any bounded term in λ. Fix $n \geq 1$. Since

$$T^n x_i \geq [P(f_1)^n x]_i + \sum_{\ell=0}^{n-1} [P(f_1)^{\ell} q(f_1)]_i,$$

and since C is a subchain of $P(f_1)$, we have

$$T^n x_i = \lambda + O(1), \quad \text{for } i \in C$$

Since j_2 cannot reach C, we have $(T^n x)_{j_2} = 0(1)$. Finally observing that $T^n v = 0(1)$, we have $\|T^n x - T^n v\|_d = \lambda + 0(1)$ whereas $\|x - v^*\|_d = \lambda + 0(1)$ as well. Conclude that for all n = 1,2,...

$$\sup_{x \in \tilde{W}} f_n(x) \geq \lim_{\lambda\to\infty} \frac{\|T^n x - T^n v\|_d}{\|x-v\|_d} = \lim_{\lambda\to\infty} \frac{\lambda+0(1)}{\lambda+0(1)} = 1$$

thus contradicting the prerequisite.

(b) Assume to the contrary that a sequence $\{x^\alpha\}_{\alpha=1}^{\infty}$ exists with $x^\alpha \in W^*$ and

(7.4) $$\lim_{\alpha\to\infty} f_m(x^\alpha) = 1 \quad \text{for some } m \geq N^2+2.$$

Due to part (b) of lemma 5.1 we have $f_n(x^\alpha) = \|T^n x^\alpha - T^n v\|_d / \|x^\alpha - v\|_d$. Hence for each $\alpha = 1,2,\ldots$ $f_n(x^\alpha)$ is unchanged by adding a multiple of $\underline{1}$ to each x^α. For the sake of notational simplicity we do this in such a way that:

(7.5) $$x^\alpha - v \geq 0 \quad \text{and} \quad (x^\alpha - v)_{min} = 0.$$

We next restrict ourselves to a subsequence of $\{x^\alpha\}_{\alpha=1}^{\infty}$ such that the same m-step policy $\xi=(f_1,\ldots,f_m)$ with $f_1,\ldots,f_m \in S_P$, achieves $T^n(x^\alpha)$ for all x^α in the subsequence and all $n \leq m$, i.e.

(7.6) $$T^n x^\alpha = \tilde{q}_n + \tilde{P}_n x^\alpha \quad \text{for all } x^\alpha \text{ in the subsequence and } n \leq m,$$

$$\tilde{q}_n = q(f_n)+P(f_n)q(f_{n-1}) +\ldots+ P(f_n)\ldots P(f_2)q(f_1)$$
$$\tilde{P}_n = P(f_n)\ldots P(f_1)$$

Observe that the existence of this subsequence is guaranteed by the fact that there is only a finite number of m-step policies.
Using lemma 5.1 part (f), (7.4) implies that $\{x^\alpha\}_{\alpha=1}^{\infty}$ is unbounded; hence it follows from (7.5) that $\lim_{\alpha\to\infty} \|x^\alpha - v\|_d = \lim_{\alpha\to\infty} (x^\alpha - v)_{max} = \infty$
Next define:

(7.7) $$y^{\alpha} = \frac{x^{\alpha}-v}{\|x^{\alpha}-v\|_d} = \frac{x^{\alpha}-v}{(x^{\alpha}-v)_{max}} .$$

Observe $0 \le y_i^{\alpha} \le 1$ for all $i \in \Omega$ and $\|y^{\alpha}\|_d = 1$. Since $\{y^{\alpha}\}_{\alpha=1}^{\infty}$, is bounded we henceforth restrict ourselves to a further subsequence which has $\lim_{\alpha\to\infty} y^{\alpha} = y^*$ (say). It then follows from (7.4) that:

$$1 = \lim_{\alpha\to\infty} f_n(x^{\alpha}) = \lim_{\alpha\to\infty} \|\tilde{q}_n + \tilde{P}_n x^{\alpha} - T^n v\|_d/(x^{\alpha}-v)_{max} =$$

$$= \lim_{\alpha\to\infty} \frac{[\tilde{P}_n(x^{\alpha}-v)]_{max} - [\tilde{P}_n(x^{\alpha}-v)]_{min} + 0(1)}{(x^{\alpha}-v)_{max}}$$

$$= [\tilde{P}_n y^*]_{max} - [\tilde{P}_n y^*]_{min} \text{ for all } n \le m.$$

Since $0 \le y^* \le 1$ for all $i \in \Omega$ this implies that

(7.8) $$[\tilde{P}_n y^*]_{max} = 1; \; [\tilde{P}_n y^*]_{min} = 0 \quad \text{for all } n \le m.$$

Recalling (7.6) we obtain:

$$T^n(x^{\alpha}) = \tilde{q}_n + \tilde{P}_n x^{\alpha} = \max_{(h_1,\ldots,h_n)} \{q(h_n)+P(h_n)q(h_{n-1})+\ldots+ P(h_n)\ldots P(h_2)q(h_1)+P(h_n)\ldots P(h_1)x^{\alpha}\}$$

Dividing this equality by $(x^{\alpha}-v^*)_{max}$, and letting α tend to infinity, we obtain:

(7.9) $$[\tilde{P}_n y^*]_i = \max_{(h_1,\ldots,h_n)} [P(h_n)\ldots P(h_1)y^*]_i, \quad \text{for all } i \in \Omega, \quad 1 \le n \le m.$$

We shall prove that

(7.10) $$\tilde{P}_n y^*_i = [P(f_n)\ldots P(f_1)y^*]_i = 0 \quad \text{for all } i \in R^* \text{ and } n = 0,1,\ldots,2N.$$

Assume to the contrary that there exists a state $j_0 \in R^*$ such that $[P(f_n)\dots P(f_1)y^*]_{j_0} = 0$ for some $n \le 2N$. Fix $f^* \in S_{RMG}$ such that R^* is the single subchain of $P(f^*)$ and recall from th.5.2 that $P(f^*)^{N^2-2N+2}_{ti} > 0$ for all $t \in \Omega$, and $i \in R^*$. Then using (7.9):

$$[\tilde{P}_n y^*]_i \ge [P(f^*)^{m-n} P(f_n)\dots P(f_1)y^*]_i \ge$$

$$P(f^*)^{m+n}_{ij_0} \; [P(f_n)\dots P(f_1)y^*]_{j_0} > 0$$

for all $i \in \Omega$ contradicting (7.8).
Define $S_n = \{i \mid \tilde{P}_n y^*_i = y^*_{max} = 1\}$. It follows from (7.8) that S_n is non-empty for $n \le m$. Using the same arguments as were used in the proof of th.5.2 with respect to the sets X(n), we obtain that there is a $k \le N - 1$ such that $S(k) \subseteq S = \cup^N_{\ell=k+1} S(\ell)$ with S being a closed subset, i.e. containing a subchain of some policy. In other words, $\hat{R}$ intersects S(r) for some $r(k \le r \le N)$.
Finally, let f be a policy that has $\hat{R}$ as its single subchain. Fix $i \in R^* \subseteq \hat{R}$; since all states in $\hat{R}$ communicate with each other under $P(\hat{f})$ there exists an integer $t \le N$ such that $\sum_{j \in S(r)} P(\hat{f})^t_{ij} > 0$. Hence $[\tilde{P}_{t+r} y^*]_i \ge \sum_{j \in S(r)} P(\hat{f})^t_{ij} \, [\tilde{P}_r y^*]_j > 0$, thus contradicting (7.10) since $t + r \le 2N$. □

We conclude this section by observing that under (H1), a number of equivalent formulations for (H2) can be obtained, e.g.:

(7.11) No policy $f \in S_P$ has a subchain within $\Omega \backslash R^*$ which cannot be reached from R^*, i.e. if S is a subchain of some policy f^0, with $S \subseteq \Omega \backslash R^*$ then there exists a policy h such that $\sum_{j \in S} P(h)^n_{ij} > 0$ for some $n \le N$.

or

(7.12) $\hat{R}$ is a *communicating system* (cf. BATHER [1]).

We refer to [6] for the proofs of these equivalences and for a more detailed investigation of the underlying structure. Note that the combination of (H1) and (H2) is trivially satisfied in the unichain case.

Observe finally that in example 3, $\hat{R} = \{1,2\}$ and that no policy has $\hat{R}$ as its single subchain.

REFERENCES:

[1] BATHER, J., *Optimal decision procedures for finite Markov chains*, Adv. in Appl. Prob. 5 (1973), parts I and II, 328-339, 521-540.

[2] BELLMAN, R., *A Markov decision process*, J. Math. Mech. 6 (1957), 679-684.

[3] 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.

[4] DENARDO, E., *Contraction Mappings in the theory underlying dynamic programming*, SIAM Rev. 9 (1967) 165-177.

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

[6] FEDERGRUEN, A. & P.J. SCHWEITZER. *A Lyapunov function for Markov Renewal Programming*, (in preparation).

[7] ____ , ____ , & H.C. TIJMS, *Contraction Mappings underlying undiscounted Markov Decision Problems*, Math. Centre Report BW 72/77 (1977) (to appear in J. Math. Anal. Appl.).

[8] HOLLADAY, J, & R. VARGA, *On powers of non-negative matrices*, Proc. Amer. Math. Soc. 9 (1958), 631-634.

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

[10] LANERY, E., *Etude Asymptotique des Systèmes Markoviens à Commande*, Rev. Inf. Rech. Op. 1 (1967), 3-56.

[11] MORTON, T, & W. WECKER, *Ergordicity and convergence for Markov Decision Processes*, Man. Sci. 23 (1977), 890-900.

[12] ODONI, A., *On finding the maximal gain for Markov Decision Processes*, O.R. 17 (1969), 857-860.

[13] PAZ, A., *Introduction to Probabilistic Automata*, Academic Press, New York (1971).

[14] SCHWEITZER, P.J., *Perturbation theory and Markovian Decision Processes*, Ph. D. dissertation MIT (1965) (MITORC report H 15).

[15] ________ , *A turnpike theorem for undiscounted Markovian Decision Processes*, presented at ORSA/TIMS, national meeting, may 1968.

[16] ________ , *Iterative solution of the functional equations for undiscounted Markov Renewal Programming*, J.M.A.A. (1971), 495-501.

[17] ________ , & A. FEDERGRUEN, *Functional Equations of undiscounted Markov Renewal Programming*, Math. Center Reports BW 60/76, BW 71/77 (1976), (to appear in Math. of. O.R.).

[18] ________ & ________ , *The asymptotic behaviour of undiscounted value-iteration in Markov Decision Problems*, Math. Center Report BW 44/76 (1976) (to appear in Math. of. O.R.).

[19] SENETA, E., *Nonnegative matrices*, Allen & Unwin, London (1973).

[20] SU, Y. & R. DEININGER, *Generalization of White's method of Successive Approximations to Periodic Markovian Decision Processes*, O.R. 20 (1972) 318-326.

[21] TYMS, H., *An iterative method for approximating average cost optimal, (s,S) inventory policies*, Zeitschrift für O.R. 18 (1974), 215-223.

[22] WHITE, D., *Dynamic Programming, Markov Chains and the method of successive approximations*, J. of. Math. Anal. and Appl. 6 (1963) 373-376.