stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE (DEPARTMENT OF OPERATIONS RESEARCH)

BW 72/77

MAART

A. FEDERGRUEN, P.J. SCHWEITZER & H.C. TIJMS

CONTRACTION MAPPINGS UNDERLYING UNDISCOUNTED MARKOV DECISION PROBLEMS

Preprint

2e boerhaavestraat 49 amsterdam

BIBLIOTHEEK MATHEMATISCH CENTRUM
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

Contraction mappings underlying undiscounted Markov decision problems. ****

by

A. Federgruen*, P.J. Schweitzer** & H.C.Tijms***

ABSTRACT

This paper is concerned with the properties of the value-iteration operator which arises in undiscounted Markov decision problems.

We give both necessary and sufficient conditions for this operator to reduce to a contraction operator, in which case the value-iteration method exhibits a uniform geometric convergence rate.

As *necessary* conditions we obtain a number of important characterizations of the chain - and periodicity structure of the problem, and as *sufficient* conditions, we give a general "scrambling-type" recurrency condition, which encompasses a number of important special cases.

Next, we show that a data-transformation turns every unichained undiscounted Markov Renewal Program into an equivalent undiscounted Markov decision problem, in which the value-iteration operator is contracting, because it satisfies this "scrambling-type" condition. We exploit this contraction-property in order to obtain lower and upper bounds as well as variational characterizations for the fixed point of the optimality equation, as well as a test for eliminating suboptimal actions.

KEY WORDS & PHRASES: *value iteration operator, contraction mapping, geometric convergence of the successive approximation scheme, chain - and periodicity structure, scrambling type conditions, data-transformations, variational characterization, test for eliminating suboptimal actions.*

* Mathematisch Centrum, Amsterdam.
** I.B.M. Thomas J. Watson Research Center, Yorktown Heights, New York.
*** Mathematisch Centrum / Vrije Universiteit, Amsterdam.
**** This report will be submitted for publication elsewhere.

1. INTRODUCTION AND SUMMARY

This paper considers undiscounted Markov Decision Processes (MDP's) with finite state- and action spaces.
$\Omega = \{1,\ldots,N\}$ 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 $(i=1,\ldots,N)$, where $\Sigma_j P_{ij}^k = 1$.

We are concerned with the behaviour of the *value-iteration* operator Q, which is defined by:

$$(1.1) \qquad Qx_i = \max_{k \in K(i)} \{q_i^k + \Sigma_{j=1}^N P_{ij}^k x_j\}, \qquad i = 1, \ldots, N$$

Denote by Q^n the n-fold application of the operator Q:

$$Q^n x = Q(Q^{n-1}x) \qquad n = 2, 3,\ldots; \qquad Q^1 x = Qx$$

Note that $Q(x+c\underline{1}) = Qx + c\underline{1}$ for every scalar c, where $\underline{1}$ is the N-vector with all components unity. As a consequence, it is useful to consider the following equivalence relation on the N-dimensional Euclidean space E^N:

$$(1.2) \qquad x \sim y \iff \text{there exists a scalar } c \text{ such that } x = y + c\underline{1}.$$

Let $\tilde{E}^N$ be the quotient space which is generated by this equivalence relation, and note that $\tilde{E}^N$ is a (N-1 dimensional) vector space, with the conventional addition and scalar multiplication. Define, the $\|x\|_d$ by (cf.BATHER [2]):

$$\|x\|_d = x_{max} - x_{min}, \text{ where}$$

$$x_{max} = \max_i x_i \quad \text{and} \quad x_{min} = \min_i x_i$$

Note that $\|x\|_d$ is a quasi-norm on E^N, and let it be the norm on $\tilde{E}^N$. The operator Q appears e.g. in the *value-iteration equations*, which were first studied by BELLMAN [3] and HOWARD [12].

$$(1.3) \qquad v(n+1)_i = Q\, v(n)_i = Q^{n+1} v(0)_i, \qquad i=1,\ldots,N; \qquad n=1,2,\ldots$$

where 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 state j.

The asymptotic behaviour of the sequence $\{Q^n x\}_{n=1}^{\infty}$, $x \in E^N$ was studied in BELLMAN [3], BROWN [4], LANERY [4], WHITE [8], SCHWEITZER [9], [20] and others. In [3] it was shown that there exists an integer $d^* \geq 1$ (which may be calculated from the periodicity and chain-structure of the problem), such that

$$(1.4) \qquad \lim_{n\to\infty} Q^{nJ+r}x - (nJ+r)g^* \quad \text{exists for } \textit{all}\ x \in E^N,$$

if and only if J is a multiple of d^*, where g^* has to be taken as the maximal gain rate vector. In addition, it was shown in [4] that whenever $\lim_{n\to\infty} Q^{nJ+r}x - (nJ+r)g^*$ exists for some particular $x \in E^N$, $J = 1,2,\ldots$ and $r = 0,\ldots,J-1$ the approach to the limit $v^*(x)$ is *geometric* i.e. there exist scalars $K = K(x)$ and $\lambda = \lambda(x)$ with $0 \leq \lambda < 1$ such that:

$$(1.5) \qquad \left|Q^{nJ+r}x - (nJ+r)g^* - v^*\right| \leq K\lambda^n, \qquad n=1,2,\ldots$$

where (g^*, v^*) satisfy the average return optimality equations:

$$(1.6) \qquad g_i^* = \max_{k \in K(i)} \Sigma_j P_{ij}^k g_j^* \qquad i=1,\ldots,N$$

$$(1.7) \qquad v_i^* + g_i^* = T\, v_i^*, \qquad i=1,\ldots,N$$

with

$$(1.8) \qquad Tx_i = \max_{k \in L(i)} \{q_i^k + \Sigma_j P_{ij}^k x_j\}, \qquad i \in \Omega \text{ and}$$

$$L(i) = \{k \in K(i) \mid g_i^* = \Sigma_j P_{ij}^k g_j^*\}, \qquad i \in \Omega.$$

The geometric convergence result in (1.5) is surprising since, example 1 below shows that, even when $d^* = 1$, the Q-operator in *general* is not a

(J-step) contraction operator (for any J = 1,2,...) nor does it ultimately reduce to such a mapping. We define the latter as in DENARDO [6], i.e.:

(1.9) Let X be a normed vector space; an operator A: $X \to X$ is a J-step contraction operator, if and only if there exists a scalar ρ, $0 < \rho \leq 1$ such that for all x, $y \in X$: $|A^J x - A^J y| \leq (1-\rho)|x-y|$, where $|\ \ |$ is the norm on X.

This contrasts with what is known to be the case (cf. DENARDO [6]) in the substochastic case where $\Sigma P_{ij}^k < 1$ $(i \in \Omega, k \in K(i))$.

The fact whether an operator A, as defined in (1.9) is J-step contracting for some J = 1,2,... is *independent of the norm* chosen on X as may easily be verified using the fact that any two norms $|x|$ and $|x|'$ are equivalent in the sense that there exists constants K and K' such that $|x| \leq K|x|'$ and $|x|' \leq K'|x|$ for all $x \in E^N$ (cf. COLLATZ [5], § 9.2).

EXAMPLE 1.

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

$g^* = [0,0]$, hence $K(i) = L(i)$ for all $i \in \Omega$.

Note that $d^* = 1$, in view of every policy being aperiodic (cf. th. 3.1 part (c) of [23]). Take x = [0,X] and y = 0. Note that,

$$T^n x = [0, \max(0, X-n)] \text{ and } T^n y = 0 \quad \text{for} \quad n = 0,1,2,\ldots \quad \text{i.e.}$$

$$1 \geq \sup\left\{\frac{\|T^n u - T^n v\|_d}{\|u - v\|_d} \;\middle|\; u,v \in E^N,\ \|u - v\|_d > 0\right\} \geq$$

$$\geq \lim_{X\to\infty} \frac{\|T^n x - T^n y\|_d}{\|x - y\|_d} = \lim_{X\to\infty} \frac{\max(0,X-n)}{X} = 1 \quad \text{for } \textit{all } n = 1,2\ldots$$

(cf. also section 7 of [24])

In this paper we give (both necessary and sufficient) conditions for the Q-operator to be a J-step contraction mapping for some J = 1,2,... The identification of these conditions is of particular importance since with Q being contracting, the geometric convergence result in (4.5) is straight-

forward (cf. theorem 1), and in addition the contraction-property may be exploited in order to obtain:

(1) a lower bound for the convergence rate of the value iteration method.

(2) upper and lower bounds, as well as variational characterizations for the fixed point v^* of the functional equation (1.7) which in this case is unique up to a multiple of $\underline{1}$ (i.e. its representation in $\tilde{E}^N$ is unique).

(3) a test for eliminating suboptimal actions in the value-iteration method.

As necessary conditions we obtain some important characterizations with respect to the chain- and periodicity structure of the problem. In addition we present a general *sufficient* condition of a "scrambling" type (cf.[1], [9]) which encompasses a number of important and easily checkable conditions. We note that in [6] a special case of this "scrambling-type" condition was used to prove the convergence of the relative cost differences.

The above results are obtained after giving the notation and preliminaries in section 2.

In [21] a data-transformation was introduced which turns every undiscounted Markov Renewal Program (MRP)(cf.[7],[3]) into an undiscounted MDP which is equivalent in the sense that it has the same maximal gain rate vector, and the same set of maximal gain policies. In addition, the transformed problem has every policy aperiodic such that the (geometric) convergence of $\{Q^n x - ng^*\}_{n=1}^{\infty}$ is guaranteed for all $x \in E^N$, i.e. $d^* = 1$ (cf.(1.4)).

In section 4, we show that for unichained MRPs, this data-transformation has the considerably stronger property of turning the MRP into an equivalent MDP, in which the Q-operator is a least N-step contracting with all of the nice consequences, mentioned above. These results are obtained by showing that the transformed problem satisfies the above "scrambling-type" condition.

2. NOTATIONS AND PRELIMINARIES

A (stationary) randomized policy f is a tableau $[f_{ik}]$ satisfying $f_{ik} \geq 0$ and $\Sigma_{k \in K(i)} f_{ik} = 1$, where f_{ik} is 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 set of all pure (non-randomized) policies (i.e. each for $f \in S_P$ $f_{ik} = 0$ or 1). Associated with each $f \in S_R$ are a

N-component reward vector q(f) and N × N matrix P(f) with

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

Note that P(f) is a stochastic matrix ($P(f)_{ij} \ge 0$; $\Sigma_{j=1}^{N} P(f)_{ij} = 1$; $1 \le i, j \le N$). For each $f \in S_R$, we define the *gain-rate vector* g(f) by:

$$(2.2) \qquad g(f) = \lim_{n\to\infty} \frac{1}{n+1} \Sigma_{\ell=0}^{n} P(f)^{\ell} q(f)$$

such that $g(f)_i$ denotes the long run average expected return per unit time, when the initial state is i, and policy f is used. We next define the *maximal gain rate* vactor g^* by:

$$(2.3) \qquad g_i^* = \sup_{f\in S_R} g(f)_i; \qquad i=1,\ldots,N.$$

Since we know from DERMAN [8] that there exists a pure policy which attains the N suprema in (2.3) simultaneously, we can define:

$$(2.4) \qquad S_{PMG} = \{f\in S_P \mid g(f) = g^*\}; \quad S_{RMG} = \{f\in S_R \mid g(f) = g^*\}$$

as the set of all pure and the set of all randomized maximal gain policies. For each policy $f\in S_R$, let R(f) denote the set of states that are recurrent under P(f). Next, define R^* as the set of states that are recurrent under some maximal gain policy.

$$(2.5) \qquad R^* = \{i\in\Omega \mid i\in R(f) \quad \text{for some } f\in S_{RMG}\} =$$
$$= \{i\in\Omega \mid i\in R(f) \quad \text{for some } f\in S_{PMG}\}$$

where the second equality in (2.5) was shown in th. 3.2 part (a) of [21]. Likewise, we define $\hat{R}$ as the set of states that are recurrent under some (arbitrary) policy

$$(2.6) \qquad \hat{R} = \{i\in\Omega \mid i\in R(f) \quad \text{for some } f\in S_R\} = \{i\in\Omega \mid i\in R(f) \quad \text{for some } f\in S_P\}$$

where the second equality is a special case of the second equality in (2.5) by taking every $q_i^k = 0$. Note that $R^* \subseteq \hat{R}$.

We next observe that there always exists a solution pair (g,v) to the optimality equations (1.6) and (1.7). In addition each pair (g,v) has $g = g^*$ -sothat the sets L(i), $i \in \Omega$, are unique-, whereas the v-part of the solution pair is *not* uniquely determined (note e.g. that if v satisfies (1.7) then so does $v + c\underline{1}$, for any scalar c). We therefore define:

$$V = \{v \in E^N \mid (g^*, v) \text{ satisfy (1.6) and (1.7)}\}.$$

We finally recall the following basic properties of the Q-operator:

$$(2.7) \qquad (x-y)_{min} \leq (Qx-Qy)_{min} \leq (Qx-Qy)_{max} \leq (x-y)_{max}$$

$$\|Qx-Qy\|_d \leq \|x-y\|_d \qquad x,y \in E^N,$$

The proof of (2.7) is easy and may be found in lemma 2.1 of [2]. The T-operator, being a special case of the Q-operator, has the same properties, and in addition:

$$(2.8) \qquad T(x+cg^*) = Tx + cg^*, \qquad \text{for all scalars c; } x \in E^N$$

which is immediate from the definition of the sets L(i).

3. NECESSARY AND SUFFICIENT CONDITIONS FOR Q BEING A (J-STEP) CONTRACTION MAPPING, AND SOME OF ITS IMPLICATIONS

Before studying necessary and sufficient conditions for Q to be a J-step contraction mapping for some J = 1,2,..., we first show that the geometric convergence of the sequence $\{Q^n x - ng^*\}_{n=1}^{\infty}$ for all $x \in E^N$, is straightforward when Q^J is a contraction mapping. We first formulate and prove this result with respect to the T-operator (cf. (1.8)). The corresponding property for the Q-operator then follows from corollary 3 below.

THEOREM 1. (Geometric convergence of value-iteration)

Let T be a J-step contraction operator on $\tilde{E}^N$, for some J = 1,2,... and some contraction factor $0 < \rho \leq 1$ (cf.(1.9)). Then, for all $x \in E^N$,

there exists a $v^* = v^*(x) \in V$ such that for all $i \in \Omega$,

$$(3.1) \qquad |T^{nJ+r}x_i - (nJ+r)g_i^* - v_i^*| \leq (1-\rho)^n \|x-v^*\|_d; \quad n=1,2,\ldots; \; r=0,\ldots J-1.$$

PROOF. Fix $x \in E^N$, and $v \in V$. Let $b(v)_i^k = q_i^k - g_i^* + \Sigma_j P_{ij}^k v_j - v_i$, and note from (1.7) that $\max_{k\in L(i)} b(v)_i^k = 0$. Define $e(n,x) = T^n x - ng^* - v = T^n x - T^n v$, where the second equality follows from a repeated application of (1.7) and (2.8). Observe next that

$$(3.2) \qquad \|e(nJ+r,x)\|_d = \|T^{nJ+r}x - T^{nJ+r}v\|_d \leq \|T^{nJ}x - T^{nJ}v\|_d \leq (1-\rho)^n \|x-v\|_d,$$

$$\text{for all } n = 1,2,\ldots; \quad \text{and} \quad r = 0,\ldots,J-1$$

where the first inequality follows from (2.7) and the second one from (1.9). Conclude that

$$(3.3) \qquad \lim_{n\to\infty} \|e(n,x)\|_d = 0$$

Next substract $(n+1)g^* - v$ from both sides of the equality:

$$T^{n+1}x_i = \max_{k\in L(i)}\{q_i^k + \Sigma_j P_{ij}^k (T^n x)_j\},$$

and use (2.9) in order to get:

$$e(n+1,x)_i = \max_{k\in L(i)}\{b(v)_i^k + \Sigma_j P_{ij}^k e(n,x)_j\} =$$

$$= e(n,x)_{min} + \max_{k\in L(i)}\{b(v)_i^k + \Sigma_j P_{ij}^k [e(n,x)_j - e(n,x)_{min}]\}.$$

In view of (3.3) it follows that for n sufficiently large only alternatives $k \in L(i)$ with $b(v)_i^k = 0$ can attain the above maxima, i.e. for all n sufficiently large we have

$$e(n+1,x)_i = \max\{\Sigma_j P_{ij}^k e(n,x)_j \mid k\in L(i) \quad \text{with} \quad b(v)_i^k = 0\}. \text{ Hence,}$$

$$(3.4) \qquad e(n,x)_{min} \leq e(n+1,x)_{min} \leq e(n+1,x)_{max} \leq e(n,x)_{max}.$$

We conclude that $\{e(n,x)_{max}\}_{n=1}^{\infty}[\{e(n,x)_{min}\}_{n=1}^{\infty}]$ decreases [increases] monotonously to a limit $\lambda^+(x)[\lambda^-(x)]$. However in view of (3.2), we have that

$$\lambda^+(x)\,\lambda^-(x) = \lim_{n\to\infty} e(n,x)_{max} - \lim_{n\to\infty} e(n,x)_{min} = \lim_{n\to\infty} \|e(n,x)\|_d = 0$$

Hence

$$\lambda^+(x) = \lambda^-(x) = \lambda(x) \quad \text{and} \quad \lim_{n\to\infty} e(n,x) = \lambda(x)\underline{1},$$

or

$$\lim_{n\to\infty} T^n x - ng^* = v^* \quad \text{where} \quad v^* = v + \lambda(x)\underline{1} \in V,$$

which proves the first assertion. This together with (3.4) lead to:

$$[T^{nJ+r}x - (nJ+r)g^* - v^*]_{min} = e(nJ+r,x)_{min} - \lambda(x) \le 0 \le$$

$$\le e(nJ+r,x)_{max} - \lambda(x) = [T^{nJ+r}x - (nJ+r)g^* - v^*]_{max},$$

so that in view of (3.2):

$$|T^{nJ+r}x_i - (nJ+r)g_i^* - v_i^*| \le \|e(nJ+r,x)\|_d \le (1-\rho)^n \|x-v^*\|_d,$$

for all $n = 1,2,\ldots$; $r = 0,\ldots,J-1$ and $i \in \Omega$.

□.

We next introduce two conditions with respect to the chain- and periodicity structure, both of which appear as *necessary* conditions for Q^J or T^J to be a contraction operator (for some $J = 1,2\ldots$).

A_1: There exists a *randomized aperiodic* policy $f \in S_{RMG}$, which has R^* as its *single* subchain.

A_2: There exists a *randomized aperiodic* policy $f \in S_R$, which has $\hat{R}$ as its *single* subchain.

The following statements are equivalent formulations for both A1 and A_2, which are expressed in terms of the structure of the *finite* set of

pure (maximal gain) policies only (cf. corollary 3.3 in [22] and th. 3.1 part (c) in [23], and observe that S_R appears as the set of all maximal gain policies, when taking $q_i^k = 0$):

A_1': Let $C^* = \{C \subseteq \Omega \mid C$ is a subchain for $P(f)$, for some $f \in S_{PMG}\}$
Then (a) for any pair $C, C' \in C^*$, there exists $\{C^{(1)} = C, C^{(2)}, \ldots, C^{(n)} = C'\}$ with $C^{(i)} \in C^*$ and $C^{(i)} \cap C^{(i+1)} \neq \emptyset$ $(i=1,\ldots,n-1)$

(b) the integers which appear as the period of some subchain of some policy in S_{PMG}, are relatively prime.

A_2': Let $\hat{C} = \{C \subseteq \Omega \mid C$ is a subchain for $P(f)$, for some $f \in S_P\}$
Then (a) for any pair $C, C' \in \hat{C}$, there exists $\{C^{(1)} = C, C^{(2)}, \ldots, C^{(n)} = C'\}$ with $C^{(i)} \in \hat{C}$ and $C^{(i)} \cap C^{(i+1)} \neq \emptyset$ $(i=1,\ldots,n)$

(b) the integers which appear as the period of some subchain of some policy in S_P, are relatively prime.

We note that whereas part (b) of A_1' implies part (b) of A_2' the parts (a) of A_1' and A_2' are mutually independent. In addition, we remark that more *efficient* procedures have been established to verify A_1 and A_2 (or alternatively A_1' and A_2'). (cf. [22] and [23]).

THEOREM 2. *(Necessary conditions for* T *to be a contraction mapping).* *Let* T *be a J-step contraction mapping on* $\tilde{E}^N$ *for some* J = 1,2,... *(cf.(1.8)). Then,*

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

(2) $g_i^* = g^*$ *for all* $i \in \Omega$; *hence* $L(i) = K(i)$, *for all* $i \in \Omega$, *and* $Qx = Tx$ *for all* $x \in E^N$.

(3) A_1 *and* A_2 *hold.*

PROOF. Let $v^*, v^{**} \in V$. By a repeated application of (1.7), we obtain, using (2.9):

$$T^J v^* = v^* + Jg^* \quad \text{and} \quad T^J v^{**} = v^{**} + Jg^*.$$

Hence,

$$\|v^* - v^{**}\|_d = \|T^J v^* - T^J v^{**}\|_d \leq (1-\rho)\|v^* - v^{**}\|_d,$$

which implies $\|v^*-v^{**}\|_d = 0$, or condition (1).

Condition (1) in turn, is equivalent with the existence of a policy $f \in S_{RMG}$, which has R^* as its *single* subchain (cf. remark 3 and th. 3.2 part (c) in [22]).

Condition A_1, i.e. the fact that even *aperiodic* policies can be found with this property, then follows from the convergence of $\{T^n x - ng^*\}_{n=1}^{\infty}$ for all $x \in E^N$ (cf. theorem 1), using th. 5.4 part (b) and th. 3.1 part (f) of [23]. The existence of a *unichained* maximal gain policy in turn implies condition (2).

Next, assume to the contrary that A_2 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

(3.5) 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 possibilities leading to a contradiction in view of the definition of $\hat{R}$, and our assumption that A_2 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 $x_i = 0(1)$ otherwise where $0(1)$ denotes any bounded term in λ. Let v^* satisfy (1.7). Since

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

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

$$(T^J x)_i = \lambda + 0(1), \quad \text{for } i \in C$$

Since j_2 cannot reach C, we have $(T^J x)_{j_2} = 0(1)$. Finally observing that $T^J v^* = 0(1)$, we have

$$\|T^J x - T^J v^*\|_d = \lambda + 0(1),$$

whereas

$$\|x - v^*\|_d = \lambda + 0(1)$$

as well. Conclude that

$$1 \geq \sup\{\frac{\|T^J u - T^J v\|_d}{\|u - v\|_d} \mid u,v \in E^N \quad \text{with} \quad \|u - v\|_d > 0\} \geq$$

$$\geq \lim_{\lambda\to\infty} \frac{\|T^J x - T^J v^*\|_d}{\|x - v^*\|_d} = 1,$$

thus contradicting the fact that T is a contraction mapping. This proves A_2 by contradiction. □

COROLLARY 3. *Fix* J = 1,2,...

(1) Q *is a* J*-step contraction operator on* $\tilde{E}^N$, *for some contraction factor* $\rho > 0$ (cf.(1.9)) *if and only if*

(2) T *is a* J*-step contraction operator on* $\tilde{E}^N$, *for some contraction factor* $\rho > 0$.

In addition both (1) *and* (2) *imply that the* Q- *and* T*-operator coincide.*

PROOF.

(2) ⟹ (1): follows from theorem 2 since condition (2) implies Q = T.

(1) ⟹ (2): we recall that the Q operator reduces to the T operator as follows:

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,... and $t \geq t_0(x)$

BIBLIOTHEEK MATHEMATISCH CENTRUM
—AMSTERDAM—

the proof of which is easy and may be found in lemma 2.2, part (g) of [24]. Next, assume to the contrary, that there exist two vectors $x,y \in E^N$, such that

$$\|T^J x - T^J y\|_d > (1-\rho)\|x-y\|_d.$$

Let $t \geq \max\{t_0(x),t_0(y)\}$ and observe, using (2.9), that

$$\|Q^J(x+tg^*) - Q^J(y+tg^*)\|_d = \|T^J(x+tg^*) - T^J(y+tg^*)\|_d =$$

$$= \|T^J x - T^J y\|_d > (1-\rho)\,\|(x+tg^*) - (y+tg^*)\|_d,$$

thus contradicting (1).

REMARK 1. If Q (or T) is a J-step contraction operator on E^N, with contraction factor ρ, then in the geometric convergence result obtained in theorem 1, an upperbound may be obtained for the number of steps J needed for contraction, i.e. there exists an integer $M \leq N^2 - 2N + 2$ and a number λ, with $0 \leq \lambda \leq (1-\rho)^{M/J}$ such that for all $x \in E^N$, there exists a $v \in V$ with:

$$|Q^{nM+r}x_i - (nM+r)g_i^* - v_i| < \lambda^n\|x-v\|_d;$$

$$n = 1,2,\ldots; \quad r = 0,\ldots,M-1; \quad i \in \Omega.$$

The upperbound on M holds whenever condition A1 is satisfied, as has been shown in [24], th. 5.2, and we know from th. 2 that A1 holds whenever Q is a (J-step) contraction operator.

In addition the upperbound on M is at least sharp up to a term of the order O(N) as has been demonstrated by example 2 in [24]. One may verify that in this example, the Q-operator is a contraction operator.

We next introduce a general "scrambling-type" recurrency condition under which the Q-operator will be shown to be a contraction operator (cf. also [1], [9]):

(S): there exists an integer $J \geq 1$, such that for every pair of J-tuples of pure policies $(f_1,\ldots,f_J)$ and $(h_1,\ldots,h_J)$:

$$(3.6) \qquad \Sigma_{j=1}^{N} \min[P(f_J)\ldots P(f_1)_{i_1 j}; \; P(h_J)\ldots P(h_1)_{i_2 j}] > 0$$

for all $i_1 \neq i_2 \in \Omega$

Theorem 4 below shows that this condition (S), encompasses a number of important and easily checkable conditions.

THEOREM 4. The following conditions are special cases of condition (S):

(1) $\Sigma_j \min(P^{k_1}_{i_1 j}, P^{k_2}_{i_2 j}) > 0$ for all $i_1 \neq i_2$ and $k_1 \in K(i_1)$, $k_2 \in K(i_2)$

(2) There exists a state s and an integer $\nu \geq 1$, such that $P(f^1)\ldots P(f^\nu)_{is} > 0$ for all $f^1, f^2, \ldots, f^\nu \in S_P$; $i \in \Omega$ (cf. White [28]).

(3) Every policy is unichained; there exists a state $s \in \Omega$ which is recurrent under every policy, and $P^k_{ss} > 0$ for all $k \in K(s)$

(4) Every policy is unichained and $P^k_{ii} > 0$ for all $i \in \Omega$, $k \in K(i)$.

PROOF. (1) ==> (S) with J = 1; (2) ==> (S) with J = ν, was shown in [28]; (3) ==> (2) with $\nu = N - 1$, was shown in [1], th. 2.

(4) ==> (S): Fix two sequences of policies $(f_N,\ldots,f_1)$ and $(h_N,\ldots,h_1)$ and $i_1, i_2 \in \Omega$ with $i_1 \neq i_2$. Let

$$S(n) = \{j \mid P(f_n)\ldots P(f_1)_{i_1 j} > 0\} \quad \text{and} \quad W(n) = \{j \mid P(h_n)\ldots P(h_1)_{i_2 j} > 0\}.$$

Note that, in view of $P_{ii}^{k} > 0$ for all $i \in \Omega$, $k \in K(i)$:

(3.7) $\qquad S(n+1) \supseteq S(n),\ W(n+1) \supseteq W(n) \qquad n = 1,2,\ldots$

Thus assuming to the contrary that $S(N) \cap W(N) = \emptyset$, it follows that $S(m) \cap W(m) = \emptyset$, for all $0 \leq m \leq N$. This in turn implies that the sequence $\{S(0) \cup W(0); ;\ldots;S(N) \cup W(N)\}$ is strictly increasing, thus leading to a contradiction: for assuming that for some $m < N$, $S(m+1) = S(m)$ and $W(m+1) = W(m)$ would imply the existence of a policy for which both $S(m)$ and $W(m)$ are closed sets of states, thus contradicting its unichainedness.

REMARK 2. Observe that condition (1) requires each $P(f)$, $f \in S_P$, to be scrambling (cf. e.g.[9]). In addition we note that conditions (1), (2) and (4) are mutually independent. To verify that (2) $\not\Rightarrow$ (1), and (2) $\not\Rightarrow$ (4), consider an example in which $S_P = \{f\}$, with

$$P(f) = \begin{vmatrix} 0 & * & 0 \\ 0 & 0 & * \\ 0 & 0 & * \end{vmatrix}$$

which satisfies (2) with $\nu = 2$ (where a $*$ indicates a positive entry). Next, the example in which $S_P = \{f_1, f_2\}$ with

$$P(f_1) = \begin{vmatrix} * & 0 & * \\ * & 0 & * \\ * & 0 & 0 \end{vmatrix} \qquad \text{and} \qquad P(f_2) = \begin{vmatrix} * & 0 & * \\ * & 0 & * \\ 0 & 0 & * \end{vmatrix}$$

satisfies (1) but not White's condition, nor (4). Finally, the example with $S_P = \{f\}$ and

$$P(f) = \begin{vmatrix} * & * & 0 \\ 0 & * & * \\ 0 & 0 & * \end{vmatrix}$$

shows (4) $\not\Rightarrow$ (1), whereas (4) $\not\Rightarrow$ (2) follows from the fact that (4) includes cases where no state is recurrent under every policy. Finally observe that condition (S) requires each policy to have a unichained and aperiodic tpm.

Theorem 5 below shows that condition (S) is sufficient for Q to be a (J-step) contraction operator:

THEOREM 5. Assume condition (S) holds for some integer $J \geq 1$. Then Q is a (J-step) contraction operator on $\tilde{E}^N$.

PROOF. The proof of this theorem is related to the one of th. 1 in [1]. First, define

$$\alpha = \min\{\sum_j \min[P(f_J) \ldots P(f_1)_{i_1 j}; P(h_J) \ldots P(h_1)_{i_2 j}] \mid \tag{3.8}$$

$$i_1, i_2 \text{ with } i_1 \neq i_2, f_k, h_k (1 \leq k \leq J)\},$$

where $\alpha > 0$ follows from (3.6) and the fact that in (3.8) the minimum is over a finite number of combinations. We shall prove that:

$$(Q^J x - Q^J y)_i - (Q^J x - Q^J y)_\ell \leq (1-\alpha)\|x-y\|_d \quad \text{for all } i, \ell \in \Omega. \tag{3.9}$$

The theorem clearly follows from (3.9). The inequality in (3.9) trivially holds when $i = \ell$. Fix now $i \neq \ell$, and let

$$Q^J x_i = q(f_J)_i + \Sigma_{k=1}^{J-1} P(f_J)\ldots P(f_{J-k+1}) q(f_{J-k})_i + P(f_J)\ldots P(f_1) x_i,$$

and

$$Q^J y = q(h_J) + \sum_{k=1}^{J-1} P(h_J)\ldots P(h_{J-k+1}) q(h_{J-k})_\ell + P(h_J)\ldots P(h_1) y_\ell$$

Next introduce the shorthand notation,

$$\beta_j = P(f_J)\ldots P(f_1)_{ij} \quad \text{and} \quad \gamma_j = P(h_J)\ldots P(h_1)_{\ell j}$$

Defining $a^+ = \max(a,0)$ and $a^- = \min(a,0)$, (with $a^+ \geq 0$, $a^- \leq 0$ and $a^+ + a^- = a$) and using the fact that

$$\Sigma_j a_j^+ = -\Sigma_j a_j^-, \quad \text{if} \quad \Sigma_j a_j = 0,$$

as well as the fact that $(a-b)^+ = a - \min(a,b)$, we obtain:

$$(Q^J x - Q^J y)_i - (Q^J x - Q^J y)_\ell \leq \Sigma_j \beta_j (x-y)_j - \Sigma_j \gamma_j (x-y)_j =$$

$$= \Sigma_j [\beta_j - \gamma_j]^+ (x_j - y_j) + \Sigma_j [\beta_j - \gamma_j]^- (x-y)_j \leq (x-y)_{max} \Sigma_j [\beta_j - \gamma_j]^+$$

$$+ (x-y)_{min} \Sigma_j [\beta_j - \gamma_j]^- = \Sigma_j [\beta_j - \gamma_j]^+ \|x-y\|_d =$$

$$= [1 - \Sigma_j \min(\beta_j, \gamma_j)] \|x-y\|_d \leq (1-\alpha) \|x-y\|_d . \qquad \square .$$

4. ON TRANSFORMING UNICHAINED MARKOV RENEWAL PROGRAMS INTO EQUIVALENT AND CONTRACTING MARKOV DECISION PROBLEMS

In this section, we consider the more general class of Markov Renewal Programs in which the times between two successive transitions of state are random variables, whose distributions depend both on the current state and the action chosen. Let $\tau_{ij}^k \geq 0$ for $i,j \in \Omega$; $k \in K(i)$ denote the *conditional* expected holding time in state i, given the action $k \in K(i)$ is chosen and that state j is the next state to be observed. We assume that the *unconditional* expected holding times:

$$T_i^k = \Sigma_j P_{ij}^k \tau_{ij}^k > 0 \quad (i \in \Omega;\ k \in K(i))$$

For each policy $f \in S_R$, $q(f)$ and $P(f)$ are defined as in section 2, whereas $g(f)_i$ denotes again the long run average return per unit time, when starting in state i. We finally recall that in this model the optimality equations (1.6) and (1.7) have to be altered as follows:

$$(4.1) \qquad g_i = \max_{k \in K(i)} \Sigma_j P_{ij}^k g_j \qquad ;i \in \Omega$$

$$(4.2) \qquad v_i = \max_{k \in L(i)} \{q_i^k - \Sigma_j P_{ij}^k \tau_{ij}^k g_j + \Sigma_j P_{ij}^k v_j\} \qquad ;i \in \Omega$$

The vector g^* and the sets S_{PMG} and S_{RMG} are defined as in section 2, where the non-emptyness of these sets in the MRP-model was shown in [13]. The properties mentioned in section 2, with respect to the set of solutions to (1.6) and (1.7) hold unaltered for (4.1) and (4.2), with the set V redefined as: $V = \{v \in E^N |$ v satisfies (4.2)$\}$. We define two undiscounted MRPs to be *equivalent* if they have the same state- and action spaces, as well as the same maximal gain rate vector and the same set of maximal gain policies.

We first recall that the gain rate vectors g(f) depend on the quantities τ_{ij}^k only through the *unconditional* holding times T_i^k. As a consequence, we conclude that every MRP is transformed into an equivalent one, by replacing $\hat{\tau}_{ij}^k = T_i^k$ $(i,j \in \Omega;\ k \in K(i))$. We thus obtain the following pair of optimality equations:

$$(4.3) \qquad g_i = \max_{k \in K(i)} \Sigma_j P_{ij}^k g_j \qquad ;i \in \Omega$$

$$(4.4) \qquad v_i = \max_{k \in L(i)} \{q_i^k - T_i^k g_i + \Sigma_j P_{ij}^k v_j\} \qquad ;i \in \Omega$$

Next, in [21] the following data-transformation was introduced which turns every MRP, with (4.3) and (4.4) as the associated pair of optimality equations into an equivalent MDP.

$$(4.5) \qquad \hat{P}_{ij}^k = (\tau/T_i^k)(P_{ij}^k - \delta_{ij}) + \delta_{ij}; \qquad i,j \in \Omega;\ k \in K(i)$$

$$\hat{q}_i^k = q_i^k/T_i^k; \qquad i \in \Omega;\ k \in K(i)$$

where $\tau > 0$ has to be chosen such that

$$(4.6) \qquad 0 < \tau \leq \min_{i,k} T_i^k/(1-P_{ii}^k)$$

so as to ensure that all $\hat{P}_{ij}^k \geq 0$ $(i,j \in \Omega;\ k \in K(i))$. Note that (4.6) is satisfied for all $0 < \tau < \min_{i,k} T_i^k$. Let V be the set of solutions to the optimality equation (4.4) and let V be the set of fixed points of the corresponding optimality equation in hte transformed MDP. Then $\hat{V} = \{v \in E^N |$ $\tau v \in V\}$, see [21]. Let $\hat{Q}$ be the value-interation operator in the transformed MDP.

Observe finally that, by taking τ strictly smaller than the upperbound in (4.6), we have all $\hat{P}^k_{ii} > 0$, which implies that every policy has an aperiodic tpm, such that for all $x \in E^N$, the geometric convergence result (1.5) holds for the $\hat{Q}$ operator, with $J = 1$, i.e. for all $x \in E^N$, there exists a vector $v \in V$, and numbers $K = K(x)$, and $\lambda = \lambda(x)$ with $0 \leq \lambda < 1$, such that:

$$(4.7) \qquad |\hat{Q}^n x - ng^* - v^*| < K\lambda^n, \qquad n = 0,1,2,\ldots$$

(To verify (4.7), cf.th. 3.1 and th. 5.1 of [23], as well as [24]).

This shows that, by applying the above data-transformation, and by subsequently doing value-iteration with respect to the transformed MDP, we find sequences which approach g^* and some $v \in V$; moreover, it follows from a generalization of Odoni [17] and from the fact that the original MRP and the transformed MDP are equivalent, that any policy which is generated by the value-iteration scheme (cf. (1.3)), for large enough n, is maximal gain.

We henceforth assume condition (H) to hold.

(H): every pure policy in the MRP is unichained.

We next make the important observation that, with τ *chosen strictly smaller than the upperbound in* (4.6), *the* Q-*operator satisfies condition* (4) *of th.*4 , *and as a consequence has the considerably stronger property of being* J-*step contracting with* $J \leq N$ (*cf. th.*5).

Note that since the $\hat{Q}$-operator is contracting under condition (H), $v \in V$ is unique up to an multiple of $\underline{1}$ (cf. th. 2), i.e. its representation v^* in $\tilde{E}^N$ is unique. In the remainder of this paper, we will show that for unichained MRP's the above data-transformation and the resulting contraction property of the operator $\hat{Q}$ in the transformed MDP may be exploited, in order to

(a) find lower and upper bounds for v^*

(b) derive variational characterizations (extremal principles) for v^*

(c) derive a test for eliminating nonoptimal actions.

We will use the following representation of $\tilde{E}^N$ (cf. section 1): $\tilde{E}^N = \{x \in E^N \mid x_N = 0\}$ such that the representation of a vector $x \in E^N$ in $\tilde{E}^N$ is given by $\tilde{x}$, with $\tilde{x}_i = x_i - x_N$, $i \in \Omega$. Note that since $\tilde{x}_{min} \leq 0 \leq \tilde{x}_{max}$,

for all $x \in E^N$.

(4.8) $\quad |\tilde{x}_i| \le \|\tilde{x}\|_d = \|x\|_d, \qquad i \in \Omega$

THEOREM 6. *Consider the* MDP *value-iteration operator* Q. *Let* Q *be a (*J*-step) contraction operator (for some* $J \ge 1$*) on* $\tilde{E}^N$, *with contraction factor* $\rho > 0$ *(cf.* (1.9)). *Define* $\tilde{Q}$ *as the reduction of the operator* Q *to* $\tilde{E}^N$, *i.e.* $\tilde{Q}: E^N \to \tilde{E}^N: x \to \tilde{Q}x = Qx - [Qx]_N \cdot \underline{1}$, *and let* v^* *be the (unique) fixed point of* Q *(or* $\tilde{Q}$*) on* $\tilde{E}^N$. *Then for all* $x \in E^N$, $n \ge 0$ *and* $0 \le r \le J - 1$.

(a) $$\tilde{Q}^{nJ+r}x_i - \rho^{-1}(1-\rho)^n\|Q^Jx-x\|_d \le v_i^* \le \tilde{Q}^{nJ+r}x_i + \rho^{-1}(1-\rho)^n\|Q^Jx-x\|_d$$

Hence,

$$\|Q^{nJ+r}x - v^*\|_d \le \rho^{-1}(1-\rho)^n\|Q^Jx-x\|_d,$$

(b) *(Alternative elimination)*
If for some $x \in E^N$, *some state* $i \in \Omega$, *and some action* $k \in K(i)$

(4.9) $$q_i^k + \Sigma_j P_{ij}^k x_j - x_i < (Q^Jx - Q^{J-1}x)_{min} - \rho^{-1}\|Q^Jx-x\|_d.$$

Then k *does not satisfy the maximum in the optimality equation* (1.7), *i.e.* k *is nonoptimal*

PROOF.

(a) Using the continuity of the $\|\ \|_d$ -norm on E^N, as well as (4.8) we obtain:

$$|\tilde{Q}^{nJ+r}x_i - v_i^*| \le \|Q^{nJ+r}x - \lim_{m\to\infty}\{Q^{mJ+r}x - (mJ+r)g^*\}\|_d =$$

$$= \lim_{m\to\infty} \|Q^{mJ+r}x - Q^{nJ+r}x\|_d = \lim_{m\to\infty} \|\Sigma_{\ell=n}^{m-1}(Q^{(\ell+1)J+r}x - Q^{\ell J+r}x)\|_d \leq$$

$$\leq \Sigma_{\ell=n}^{\infty} \|Q^{(\ell+1)J+r}x - Q^{\ell J+r}x\|_d \leq \Sigma_{\ell=n}^{\infty}(1-\rho)^{\ell} \|Q^{J+r}x - Q^{r}x\|_d \leq$$

$$\leq \rho^{-1}(1-\rho)^{n} \|Q^{J}x-x\|_d$$

where the last inequality follows from (2.8).

(b) It follows from the proof of theorem 1 of [17] that $g^* \geq (Q^Jx-Q^{J-1}x)_{\min}$. Suppose alternative $k \in K(i)$ which satisfies (4.9), attains the maximum in the optimality equation (1.7). Note from corollary 3 that the Q-operator and T-operator coincide. Then, using part (a) and the fact that $v^* \in V$, we have

$$q_i^k + \Sigma_j P_{ij}^k x_j - x_i \geq q_i^k - g^* + \Sigma_j P_{ij}^k v_j^* - v_i^* + \Sigma_j P_{ij}^k (x_j - v_j^*) -$$

$$- (x_i - v_i^*) + g^* \geq (x-v^*)_{\min} - (x-v^*)_{\max} + g^* = -\|x-v^*\|_d + g^* \geq$$

$$\geq -\rho^{-1}\|Qx-x\|_d + (Q^Jx-Q^{-1}x)_{\min}.$$

REMARK 3. The reduction of the Q-operator to E^N, was first used in White [28], in order to ensure the boundedness of his value-iteration scheme. The lower- and upper bounds for v^* are in fact generalizations of the lower- and upper bounds obtained by MAC QUEEN [15] and PORTEUS [18] for MDP's. Note that our bounds with n = 0 coincide with the analogon of Mac Queen's bounds, whereas the analogon of Porteus' bounds is obtained by taking n = 1.

By using the above data-transformation, and by applying th. 6 to the transformed MDP, we obtain upper- and lower bounds as well as variational characterizations for each of the components of v^*, and in addition a test for eliminating *non-optimal actions*.

COROLLARY 7. *Consider a unichained* MRP. *Fix* $\tau < \min_{i,k} T_i^k/(1-P_{ii}^k)$ *and let* $\hat{Q}$ *be the value-iteration operator in the transformed* MDP (*cf.*(4.5) *and* (4.6)). *Next, let* $\tilde{Q}$ *be the reduction of* $\hat{Q}$ *to* $\tilde{E}^N$, *i.e.* $\tilde{Q}x = \hat{Q}x - [Qx]_N 1$ *for all*

$x \in E^N$. *Finally, let* ρ *be the (N-step) contraction factor of the operator* $\hat{Q}$ (*cf*(1.9) and *th.* 4). *Then,*

(a) $$\tilde{Q}^{nN+r}x_i - \rho^{-1}(1-\rho)^n \|\hat{Q}^N x - x\|_d \le v_i^* \le \tilde{Q}^{nN+r}x_i + \rho^{-1}(1-\rho)^n \|Q^N x - x\|_d$$

for all $x \in E^N$, *and* $n = 0,1,\ldots;\quad r = 0,\ldots,N-1$

(b) $$v_i^* = \max_{x\in E^N} \{\tilde{Q}^{nN+r}x_i - \rho^{-1}(1-\rho)^n \|\hat{Q}^N x - x\|_d\}$$

$$= \min_{x\in E^N} \{\tilde{Q}^{nN+r}x_i + \rho^{-1}(1-\rho)^n \|\hat{Q}^N x - x\|_d\}$$

$i \in \Omega;\ n = 0,1,\ldots..;\quad r = 0,\ldots.,n-1.$

(c) *If for some* $x \in E^N$, *some state* $i \in \Omega$, *and some action* $k \in K(i)$

$$q_i^k + \Sigma_j \hat{P}_{ij}^k x_j - x_i < (\hat{Q}_x^N - \hat{Q}^{N-1}x)_{\min} - \rho^{-1}\|\hat{Q}_x^N - x\|_d,$$

then k *is nonoptimal.*

The variational characterizations in part (b) follow from part (a) by taking $x = v \in V$. Variational characterizations for g^* were recently obtained in [25]. One might use both lower and upper bounds for v^*, and the test for eliminating suboptimal actions (cf. part (a)), in the course of the following value-iteration scheme for finding g^*, v^* and some maximal gain policy.

(4.10) $$y(n)_i = \tilde{Q}y(n-1)_i = \max_{k\in K(i)}\{\hat{q}_i^k + \Sigma_j \hat{P}_{ij}^k y(n-1)_j\} +$$

$$- \max_{k\in K(N)}\{\hat{q}_N^k + \Sigma_j \hat{P}_{Nj}^k y(n-1)_j\}, \qquad i \in \Omega$$

with $y(0) \in E^N$ chosen arbitrarily.

Let f_n be a policy which achieves the N maxima in (4.10). Define

$$\theta_L(n) = [\hat{Q}y(n-1) - y(n-1)]_{\min};\ \theta_U(n) = [\hat{Q}y(n-1) - y(n-1]_{\max}.$$

The sequence $\{y(n)\}_{n=1}^{\infty}$ has the following, easily verified and previously discussed properties.

(a) $y(n) \to v^*$

(b) $\theta_L(n) \le g(f_n) \le g^* \le \theta_U(n)$ (cf. HASTINGS [6] and ODONI [17]) with

$$\lim_{n\to\infty} \theta_L(n) = g^* = \lim_{n\to\infty} \theta_U(n)$$

(c) f_n is maximal gain, for all n sufficiently large (cf. ODONI [17])

E.g. whenever at some stage n, i.e. for $x = y(n)$, the test in part (c) of cor. 7 is met for some $i \in \Omega$, and $k \in K(i)$, k may be deleted permanently from K(i) thus reducing the number of calculations in the following iterations. However, both the application for the bounds for v^* as the use of the elimination test require the computation of at least some lower bound of the contraction factor ρ, i.e. of the scrambling coefficient α, as defined in the right hand side of (3.6). Note that,

$$(4.11) \qquad 0 < \hat{\rho} = \min\{P(f_N)\ldots P(f_1)_{ij} > 0 \mid i,j \in \Omega; f_1,\ldots,f_N \in S_P\} \le \alpha \le \rho$$

where the last inequality follows form the proof of th. 5, and where the second one may be verified as follows: Let the minimum in (4.11) be attained for $s,t \in \Omega$; f_k, $h_k \in S_P (1 \le k \le N)$ and fix γ such that

$$\beta = \min[P(f_N^*)\ldots P(f_1^*)_{s\gamma};\ P(h_N^*)\ldots P(h_1^*)_{t\gamma}] > 0.$$

Then, $\alpha \ge \beta \ge \rho$. $\hat{\rho}$ may be computed as follows. Let x^o be defined by

$$x_i^o = \min\{P_{ij}^k > 0 \mid j \in \Omega,\ k \in K(i)\}, \qquad i \in \Omega.$$

Then, $\hat{\rho} = [U^N x^o]_{\min}$, where the operator U is defined by:

(4.11) $\qquad Ux_i = \min_{k \in K(i)} \Sigma_j P^k_{ij} x_j, \qquad i \in \Omega;\ x \in E^N$

Observe from the analogon of (2.7) that

$$\hat{\rho} = [U^N x^o]_{min} \geq [U^{N-1} x^o]_{min} \geq \ldots \geq x^o{}_{min}$$

such that

$$\hat{\hat{\rho}} = \min\{P^k_{ij} > 0 \mid i,j \in \Omega,\ k \in K(i)\}$$

is a lower bound of ρ (it may however be worthwhile to do a number of iterations with the U-operator on x^o, in order to obtain a better approximation of ρ).

If the employed approximation for $\rho << 1$, then the bounds of cor.7 part (a) will not be sharp, and the test of part (c) will not be met unless $\|x-v\|_d$ is very close to zero, namely when $x = y(n)$ and $n >> 1$. Hence, if $\hat{\rho} << 1$, the bounds and the test will only be important near the very end of the calculations. In addition one should observe that N represents the worst case behaviour for the number of steps needed for contraction, which is enormously high, compared with the empirical fact that in most cases $J = 1$ or 2 (cf. e.g. [26] and [27]).

Alternatively, one might want to use the test part of (c) in combination with a device, given recently by Hastings [11] in order to eliminate actions on a provisional rather than on a permanent basis.

REMARK 4. Hastings' test works as follows. Let

$$g(n,i,k) = \hat{Q}y(n-1) - \hat{q}^k_i - \Sigma_j \hat{P}^k_{ij}\, y(n-1)_j \geq 0;\ \phi(n) = \theta_U(n) - \theta_L(n),$$

and $\quad H(m,n,i,k) = g(n,i,k) - \sum_{c=n}^{m=1} \phi(c),\ m > n.$

Then, action $k \in K(i)$ is non-optimal at value iteration stage m, if $H(m,n,i,k) > 0$ (for some $n < m$).

We observe that theorem 2 of [11] holds unconditionally, for every (multichain) MDP, i.e. there is a stage after which no nonoptimal action will pass the above test. This is an immediate consequence of the geometric convergence result in (1.5)(cf. also [24]). However, whereas the *identification* of non-optimal actions is possible in the unichain case, using the above value-iteration scheme and cor. 7 part (c). this is (so far) infeasible for the general *multichain* case.

REFERENCES.

(1) ANTHONISSE J. & H. TIJMS, *On the stability of products of stochastic matrices*, Math. Center report BW 58/75 (1975), to appear in J.M.A.A.

(2) BATHER J., *Optimal decision procedures for finite Markov Chains*, Adv. in Appl. Prob. 5 (1973), p. 521-540.

(3) BELLMAN R., *A Markov Decision Process*, J. Math. Mech. 6 (1957), p. 679-684.

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

(5) COLLATZ L., *Funktional Analysis und Numerische Mathematik*, Berlin, Springer Verlag, (1964).

(6) DENARDO E., *Contraction mappings in the theory underlying Dynamic Programming*, SIAM Review 9 (1967), p. 165-177.

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

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

(9) HAJNAL J., *Weak Ergodicity in nonhomogeneous Markov chains*, Proc. Cambridge Philos. Soc. 54 (1958), p. 233-246.

(10) HASTINGS N., *Bounds on the gain of a Markov Decision Process*, Op. Res. 19 (1971), p. 240-244.

(11) _____ _., *A test for nonoptimal actions in undiscounted finite Markov Decision Chains*, Man. Sci. 23 (1976), p. 87-92.

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

(13) JEWELL W., *Markov Renewal Programming*, Op. Res. 11 (1963), p. 938-971.

(14) LANERY E., *Etude asymptotique des systèmes Markoviens à commande*, R.I.R.O. 1 (1967), p. 3-56.

(15) MACQUEEN J., *A test for suboptimal actions in Markovian Decision Processes*, Op. Res. 15 (1967), p. 559- 562.

(16) MORTON T. & W. WECKER, *Discounting, Ergodicity and Convergence for Markov Decision Processes*, (to appear in Man. Sc.).

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

(18) PORTEUS E., *Some bounds for discounted sequential decision processes*, Man. Sc. 18 (1971), p. 7-11.

(19) SCHWEITZER P.J., *Perturbation Theory and Markovian Decision Processes*, Ph. D. Dissertation, M.I.T. Operations Research Center Report 15 (1965).

(20) _____ __., *A turnpike theorem for undiscounted Markovian Decision Processes*, presented at ORSA/TIMS, National Meeting, May 1968.

(21) _____, *Iterative solution of the Functional Equations of undiscounted Markov Renewal Programming*, J.M.A.A. 34 (1971), p. 495-501.

(22) _____, & A. FEDERGRUEN, *Functional Equations of undiscounted Markov Renewal Programming*, Math. Center Report, BW 60/76, (1976) (to appear in Math. of Op. Res.).

(23) _____ & _____, *The asymtotic behaviour of undiscounted value iteration in Markov Decision Problems*, Math. Center Report BW 44/76 (1976).

(24) _____ & _____, *Geometric Convergence of value-iteraiton in multi-chains Markov Decision Problems.*

(25) _____ & _____, *Variational characterizations in Markov Renewal Programs.*

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

(27) TIJMS H., *An iterative method of approximating average cost optimal (s,S) inventory policies,* Zeitschrift für O.R. 18 (1974), p. 215-223.

(28) WHITE D., *Dynamic programming, Markov Chains, and the method of successive approximations,* J.M.A.A. 6 (1963), p. 373-376.