stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 74/77 MAART
(DEPARTMENT OF OPERATIONS RESEARCH)

A. FEDERGRUEN & H.C. TIJMS

# THE OPTIMALITY EQUATION IN AVERAGE COST DENUMERABLE STATE SEMI-MARKOV DECISION PROBLEMS, RECURRENCY CONDITIONS AND ALGORITHMS

Preprint

2e boerhaavestraat 49 amsterdam

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

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

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

# The optimality equation in average cost denumerable state semi-Markov decision problems, recurrency conditions and algorithms[*]

by

A. Federgruen & H.C. Tijms[**]

## ABSTRACT

This paper is concerned with the optimality equation for the average costs in a denumerable state semi-Markov decision model. It will be shown that under each of a number of recurrency conditions on the transition probability matrices associated with the stationary policies, the optimality equation has a bounded solution. This solution indeed yields a stationary policy which is optimal for a strong version of the average cost optimality criterion. Besides the existence of a bounded solution to the optimality equation, we will show that both the value-iteration method and the policy-iteration method can be used to determine such a solution. For the latter method we will prove that the average costs and the relative cost functions of the policies generated converge to a solution of the optimality equation.

KEY WORDS & PHRASES: *Semi-Markov decision model, denumerable state space, average costs, optimality equation, recurrency conditions, value-iteration method, policy-iteration method, convergence results.*

[*] This report will be submitted for publication elsewhere

[**] Mathematisch Centrum - Vrije Universiteit

## 1. INTRODUCTION

We consider a semi-Markov decision model specified by five objects $(I, A(i), p_{ij}(a), c(i,a), \tau(i,a))$. The state space I is denumerable and, for any $i \in I$, the set $A(i)$ denotes the set of possible actions for state i. If the system is in state i at a decision epoch and action a is chosen, then an immediate (expected) cost of $c(i,a)$ is incurred and the (expected) time until the next decision epoch is $\tau(i,a)$ where the next state will be j with probability $p_{ij}(a)$. Throughout this paper we make the following assumptions.

A1. There is a finite number M such that $|c(i,a)| \leq M$ for all $i \in I$ and $a \in A(i)$.

A2. There is finite number $\varepsilon > 0$ and a finite number M such that $\varepsilon \leq \tau(i,a) \leq M$ for all $i \in I$ and $a \in A(i)$.

A3. For any $i \in I$, the set $A(i)$ is a compact metric space such that both $c(i,a)$, $\tau(i,a)$ and $p_{ij}(a)$ for any $j \in I$ are continuous on $A(i)$.

Denote by F the class of all functions f which add to each state $i \in I$ a single action $f(i) \in A(i)$. Then $F = \times A(i)$ is a compact metric space in the product topology. For any $f \in F$, let $P(f)$ be the stochastic matrix whose (i,j)th element is $p_{ij}(f(i))$ and let $P^n(f) = (p^n_{ij}(f))$ be the n-fold matrix product of $P(f)$ with itself, $n \geq 1$. A policy $\pi$ for controlling the system is any (possibly randomized) rule for choosing actions. For any $f \in F$, denote by $f^{(\infty)}$ the stationary policy which prescribes to take action $f(i)$ whenever the system is in state i. Denote by $X_n$ and $a_n$ the state of the action chosen at the nth decision epoch for $n = 0,1,\ldots.$(the 0th decision epoch is epoch 0). For $n = 1,2,\ldots.$, let $\tau_n$ be the time the (n-1)st and the nth decision epoch. Denote by $E_\pi$ the expectation when policy $\pi$ is used.

In this paper we will be concerned with the optimality equation for the average cost case. Therefore we consider the following three recurrency conditions.

C1. There is a state $s \in I$ and a finite number B such that $E_{f(\infty)}\{N|X_0=i\} \leq B$ for all $i \in I$ and $f \in F$ where $N = \inf\{n \geq 1 | X_n = s\}$.

C2. There is a finite set $K \subset I$, an integer $\nu \geq 1$ and a number $\rho > 0$ such that

$$\Sigma_{j \in K} p_{ij}^{\nu}(f) \geq \rho \qquad \text{for all } i \in I \text{ and } f \in F.$$

Further, for any $f \in F$ the stochastic matrix $P(f)$ has no two disjoint closed sets.

C3. There is an integer $\nu \geq 1$ and a number $\rho > 0$ such that

$$\Sigma_{j \in I} \min[p_{i_1 j}^{\nu}(f),\ p_{i_2 j}^{\nu}(f)] \geq \rho \qquad \text{for all } i_1, i_2 \in I \text{ and } f \in F.$$

It is known that under condition C1 the optimality equation for the average cost case has a bounded solution, cf. [4], [5], [11], and [17] where in [11] (see chapter 5 and section 12.6) a condition was considered which is more general than C1 and even allows for unbounded costs. For the case of unbounded costs, conditions under which the optimality equation for the average costs applies, were also given in [14]. The conditions in [11] and [14] both assume the existence of a fixed regeneration state s. It may be interesting to note that a careful examination of the proofs in section 6.7 in [17] and in particular the proof of Theorem 6.19 reveals that we may somewhat weaken C1 by allowing that state s may depend on $f \in F$.

The condition C2 was first used in [11] where this condition was called the simultaneous Doeblin condition. Observe that for each $f \in F$ the stochastic matrix $P(f)$ satisfies the so-called Doeblin condition for Markov chain theory e.g. [6]. Under condition C2 the existence of an optimal stationary policy for the limsup average cost criterion was shown in [11] where also several other sufficient conditions for the existence of an average cost optimal policy were found.

The condition C3 says that for any $f \in F$ the stochastic matrix $P^{\nu}(f)$ has a positive ergodic coefficient of at least $\rho$. Clearly, under C3 we have that any $P(f)$ is aperiodic and has no two disjoint closed sets. Using a notion introduced in [9], we could call C3 a simultaneous scrambling condition, cf. also [20].

In this paper we shall give a unified proof that under each of the conditions C1, C2 and C3 the optimality equation for the average costs has a bounded solution. This will be done in section 2 and the proof will be based both on an analysis of the asymptotic behaviour of the n-step transition probability matrices $P^n(f)$ and on a simple but very useful data-transformation introduced in [19]. Also we give some interdependencies between the conditions C1, C2 and C3.

It is important to note that the existence of a bounded solution to the optimality equation implies the existence of an optimal stationary policy among the class of all policies with respect to a strong version of the average cost optimality criterion which implies essentially weaker versions usually considered in the literature, cf. [8] and Theorem 2.2 of the next section. Further we note that after having established the optimality equation for the average costs a repeated application of this result yields a sequence of optimality equations that are involved when considering the more sensitive and selective n-discounted optimality criteria, thus showing the existence of stationary n-discounted optimal policies, cf. [13].

Besides the existence of a bounded solution to the optimality equation for the average costs, we will consider the problem of determining such a solution which in its turn yields an optimal stationary policy. In section 2 we shall show that under each of the conditions C1, C2 and C3 the value-iteration method can be used to determine a bounded solution to the optimality equation. The policy-iteration method will be considered in section 4. Under condition C1 we shall prove that the average costs and the relative cost functions of the policies generated by this method converge to a solution of the optimality equation. This result considerably generalizes a related result in [4].

## 2. THE OPTIMALITY EQUATION.

In this section we shall establish the existence of a bounded solution to the optimality equation for the average costs. To do this, we first give the following results.

LEMMA 2.1. Suppose C3 holds. Then for each $f \in F$ there is a probability distribution $\{\pi_j(f),\ j \in I\}$ such that

$$(2.1)\qquad \Big|\sum_{j\in A} p^n_{ij}(f) - \sum_{j\in A} \pi_i(f)\Big| \le (1-\rho)^{[n/\nu]} \quad \text{for all } i\in I,\ A\subseteq I \text{ and } n\ge 1.$$

PROOF. The proof is a minor modification of the proof of Theorem 1 in [1].

In the next lemma we give sufficient conditions for C3.

LEMMA 2.2. Condition C3 holds under each of the following three conditions.
C3a. There is an integer $\nu\ge 1$, a number $\rho>0$ and for each $f\in F$ there is a state $s(f)$ such that $p^{\nu}_{is(f)} \ge \rho$ for all $i \in I$.

C3b. There is a number $\rho > 0$ such that $\Sigma_{j\in I}\ \min[p_{i_1j}(a_1),p_{i_2j}(a_2)] \geq \rho$ for all $i_1,i_2 \in I$ with $i_1 \neq i_2$ and all $a_1 \in A(i_1)$ and $a_2 \in A(i_2)$.

C3c. Condition C2 holds and for each $f \in F$ the stochastic matrix $P(f)$ is aperiodic.

PROOF. It is obvious that both C3a and C3b imply C3. In fact C3b is equivalent to C3 with $\nu=1$. Suppose now C3c holds. We shall now prove that C3c implies C3a and so C3. We first note that any $P(f)$ is an aperiodic stochastic matrix which satisfies the Doeblin condition from Markov chain theory and has no two disjoint closed sets. Hence for any $f \in F$ the stochastic matrix $P(f)$ has a unique stationary probability distribution $\{\pi_j(f),\ j\in I\}$ (say) such that (e.g. [6])

$$\text{(2.2)} \qquad \lim_{n\to\infty} p^n_{ij}(f) = \pi_j(f) \qquad \text{for all } i,j \in I.$$

Using $\Sigma_{j\in K}\ p^{n+\nu}_{ij}(f) = \Sigma_{k\in I}p^n_{ik}(f)\ \Sigma_{j\in K}p^{\nu}_{kj}(f) \geq \rho$ for all $n \geq 0$, it follows from (2.2) that $\Sigma_{j\in K}\pi_j(f) \geq \rho$ for all $f \in F$. Hence for each $f \in F$ there is a state $j \in K$ such that $\pi_j(f) \geq \rho/|K|$. For any $k \in K$, define now $F_k = \{f\in F\,|\,\pi_k(f) \geq \rho/|K|\}$. By Theorem 11.4 and Lemma 10.2 in [11], we have for any $j \in I$ that $\pi_j(f)$ is continuous on $F$. Using this fact it immediately follows that for any $k \in K$ the set $F_k$ is closed and hence compact. For fixed $k \in K$ and $f \in F_k$, define (cf. the proof of Lemma 11.6 in [11]),

$$n_k(f) = \inf\,\{n\geq 1\,|\,p^m_{ik}(f) > \rho/2|K| \quad \text{for all } m \geq n \text{ and } i \in K\}.$$

Then, by (2.2), $n_k(f)$ exists and is finite. Moreover, using the fact that $P^m(f)$ is continuous on $F$ for all $m \geq 1$ as easily follows by induction from assumption A3, it is immediately verified that for each $k \in K$ the set $\{f\in F_k\,|\,n_k(f) \geq \alpha\}$ is closed for any real $\alpha$. Hence for each $k \in K$, the function $n_k(f)$ is upper semi-continuous on the compact set $F_k$ and so, by Proposition 10 on p. 161 in [18] and the finiteness of K, there is an integer $\mu \geq 1$ such that $n_k(f) \leq \mu$ for all $k \in K$ and $f \in F_k$. This shows that for any $k \in K$ and $f \in F_k$ we have $p^{\mu}_{ik}(f) > \rho/2|K|$ for all $i \in K$ and so $p^{\nu+\mu}_{ik}(f) \geq \Sigma_{j\in K}\ p^{\nu}_{ij}(f)\ p^{\mu}_{jk}(f) \geq \rho^2/2|K|$ for all $i \in I$. This proves that C3a holds which completes the proof of the lemma.

By the lemmas 2.1 and 2.2 we have that each of the conditions C3a, C3b and C3c in lemma 2.2 implies the condition C2 which in its turn implies the condition

C4. There is an integer $\nu \geq 1$, a number $\rho > 0$ such that for any $P(f)$ there is a probability distribution $\{\pi_j(f),\ j \in I\}$ satisfying (2.1).

REMARK 1. Under the assumption that there is some state $i_0$(say) which is an aperiodic positive recurrent state under any $P(f)$, $f \in F$, we have the equivalencies C1 $\Leftrightarrow$ C2 $\Leftrightarrow$ C3 $\Leftrightarrow$ C4. Although we shall not need this result, we include its proof. (i) C1 $\Leftrightarrow$ C2. This equivalence was established in [11] (see section 12.6) and even holds without the assumption that $i_0$ is aperiodic. (ii) C2 $\Rightarrow$ C3 $\Rightarrow$ C4. Together C2 and the aperiodicity of state $i_0$ imply condition C3c in Lemma 2.2 and so, by the Lemmas 2.1 and 2.2 we get the desired result. (iii) C4 $\Rightarrow$ C2. Using (2.1) and the fact that $P^m(f)$ is continuous on F for all $m \geq 1$, it readily follows that for each $j \in I$ the function $\pi_j(f)$ is continuous on the compact set F. Hence, since $\pi_{i_0}(f) > 0$ for $f \in F$, we have for some number $a > 0$ that $\pi_{i_0}(f) \geq a/2$ for all $f \in F$. Together this result and (2.1) imply the existence of an integer $\mu \geq 1$ such that $p^{\mu}_{ii_0}(f) \geq a/2$ for all $i \in I$ and $f \in F$ which verifies C2.

To establish the optimality equation, we shall employ a simple but very useful data-transformation introduced in [19]. We associate with the semi-Markov model a discrete-time Markov decision model with state space I, the set A(i) as set of possible actions for state i, one-step costs $\hat{c}(i,a)$, one step transition times $\hat{\tau}(i,a) \equiv 1$ and one-step transition probabilities $\hat{p}_{ij}(a)$ where, for all $i,j \in I$ and $a \in A(i)$

$$\hat{c}(i,a) = \frac{c(i,a)}{\tau(i,a)} \quad\text{and}\quad \hat{p}_{ij}(a) = \frac{\tau}{\tau(i,a)}\{p_{ij}(a)-\delta_{ij}\} + \delta_{ij}$$

in which the Kronecker function $\delta_{ij} = 1$ for $j = i$ and $\delta_{ij} = 0$ for $j \neq i$ and $\tau$ is a fixed number such that

$$0 < \tau < \delta = \inf_{i,a}\{\tau(i,a)/(1-p_{ii}(a))\,|\,p_{ii}(a) < 1\}.$$

Observe that $\delta > 0$ and the assumptions A1 - A3 also apply to the transformed model. Further, letting the finite positive number $\gamma$ be equal to $\sup_{i,a}\tau(i,a)$, it is readily verified that for all $i \in I$ and $a \in A(i)$ we have that $\{\hat{p}_{ij}(a),\ j \in I\}$ is a probability distribution with

(2.3) $\hat{p}_{ii}(a) \geq 1 - \frac{\tau}{\delta} > 0$ and $\hat{p}_{ij}(a) \geq \frac{\tau}{\gamma}\, p_{ij}(a)$ for $j \neq i$.

By the first part of (2.3), we have that for any $f \in F$ the stochastic matrix $\hat{P}(f)$ is *aperiodic*. This aperiodicity will play a crucial role in the analysis below. Also, letting the finite positive number $\phi$ be equal to min $\lceil 1-\tau/\delta, \tau/\gamma \rceil$ and using (2.3), it is immediately verified that for any set $A \subseteq I$ and all $n \geq 1$,

(2.4) $\sum_{j\in A} \hat{p}^{n}_{ij}(f) \geq \phi^{n} \sum_{j\in A} p^{n}_{ij}(f)$ for all $i \in I$ and $f \in F$.

This inequality implies that if condition C3 holds, this condition also applies to the stochastic matrices $\hat{P}(f)$ of the transformed model. Also, it follows from (2.4) and the aperiodicity of the stochastic matrices $\hat{P}(f)$, $f \in F$ that if condition C2 holds, then condition C3c in Lemma 2.2 applies to the stochastic matrices $\hat{P}(f)$, $f \in F$. In case condition C1 holds, then, by Theorem 11.3 in [11], condition C2 holds and so condition C3c in Lemma 2.2 applies to the stochastic matrices $\hat{P}(f)$, $f \in F$. Hence, using the Lemmas 2.1 and 2.2 , we have that under each of the conditions C1, C2 and C3 there is an integer $\nu \geq 1$ and a number $\rho > 0$ such that for any $f \in F$ the stochastic matrix $\hat{P}(f)$ has a stationary probability distribution $\{\hat{\pi}_j(f),\ f \in F\}$ for which

(2.5) $\left| \sum_{j\in A} \hat{p}^{n}_{ij}(f) - \sum_{j\in A} \hat{\pi}_j(f) \right| \leq (1-\rho)^{[n/\nu]}$ for all $i\in I$, $A\subseteq I$ and $n\geq 1$.

This result will underly the derivation of the optimality equation for the transformed model from which we easily get the optimality equation for the semi-Markov decision model considered. Before showing this, we give the following lemma.

LEMMA 2.3. Let $\{h_n(.),\ n \geq 1\}$ be a sequence of bounded functions on I such that, for some bounded function h(.) on I, $\lim_{n\to\infty} h_n(i) = h(i)$ for all $i \in I$. Then, for any $i \in I$,

$$\lim_{n\to\infty} \min_{a\in A(i)} \{c(i,a) + \sum_{j\in I} p_{ij}(a)\, h_n(j)\} = \min_{a\in A(i)} \{c(i,a) + \sum_{j\in I} p_{ij}(a)h(j)\}.$$

PROOF. Fix $i \in I$. For any $n \geq 1$, let action $a_n$ mimimize $c(i,a) + \Sigma_j p_{ij}(a)h_n(j)$ for $a \in A(i)$. Observe that, by A3, such a minimizing action exists.

Now, let $\{n_k, k \geq 1\}$ be any infinite sequence of positive integers. Since $A(i)$ is a compact metric space, we can choose an action $a^* \in A(i)$ and a subsequence $\{t_k, k \geq 1\}$ of $\{n_k, k \geq 1\}$ such that $a_{t_k} \to a^*$ as $k \to \infty$. Using the fact that, by A3, $\Sigma_{j\in A}\, p_{ij}(a_{t_k}) \to \Sigma_{j\in A} p_{ij}(a^*)$ as $k \to \infty$ for any set $A \subseteq I$ and using Proposition 18 on p. 232 in [18], it follows from

$$c(i,a_{t_k}) + \sum_{j\in I} p_{ij}(a_{t_k})h_{t_k}(j) \leq c(i,a) + \sum_{j\in I} p_{ij}(a)h_{t_k}(j)$$

for all $a \in A(i)$ and $k \geq 1$, that

$$\lim_{k\to\infty} \min_{a\in A(i)} \{c(i,a) + \sum_{j\in I} p_{ij}(a)h_{t_k}(j)\} = \min_{a\in A(i)} \{c(i,a) + \sum_{j\in I} p_{ij}(a)h(j)\}$$

which proves the lemma.

We now prove the main result of this section.

THEOREM 2.1. Under each of the conditions C1, C2 and C3 there exists a finite constant $g^*$ and a bounded function $v^*(i)$, $i \in I$ such that

$$\text{(2.6)} \qquad v^*(i) = \min_{a\in A(i)} \{c(i,a) - g^*\tau(i,a) + \sum_{j\in I} p_{ij}(a)v^*(j)\} \text{ for all } i \in I.$$

The constant $g^*$is uniquely determined and the bounded function $v^*(i)$, $i \in I$ is uniquely determined up to an additive constant.

PROOF. Consider first the transformed model. As shown above, there is an integer $\nu \geq 1$ and a number $\rho > 0$ such that for any $f \in F$ the stochastic matrix $\hat{P}(f)$ has a stationary probability distribution satisfying (2.5). To verify the optimality equation for the transformed model, consider first the discounted cost criterion. For any $0 < \beta < 1$, define for each policy $\pi$ (observe that $\hat{c}(i,a)$ is uniformly bounded in $i,a$),

$$\hat{V}_\beta(i,\pi) = E_\pi[\sum_{n=0}^{\infty} \beta^n \hat{c}(X_n,a_n)|X_0=i] \qquad \text{for } i \in I,$$

and let $\hat{V}_\beta(i) = \inf_\pi \hat{V}_\beta(i,\pi)$, $i \in I$. It is known that for any $0 < \beta < 1$ the function $\hat{V}_\beta(i)$, $i \in I$ is the unique bounded solution to (e.g. [15])

$$\text{(2.7)} \qquad \hat{V}_\beta(i) = \min_{a\in A(i)} \{\hat{c}(i,a) + \beta \sum_{j\in I} \hat{p}_{ij}(a)\, \hat{V}_\beta(j)\}, \qquad i \in I,$$

and, moreover,

$$(2.8) \qquad \hat{V}_\beta(i,f_\beta^{(\infty)}) = \hat{V}_\beta(i) \quad \text{for all } i \in I,$$

for any $f_\beta \in F$ such that $f_\beta(i)$ minimizes the right-side of (2.7) for all $i \in I$. For any $0 < \beta < 1$ and $f \in F$, we have

$$(2.9) \qquad \hat{V}_\beta(i,f^{(\infty)}) = \sum_{n=0}^{\infty} \beta^n \sum_{j\in I} \hat{p}_{ij}^n (f)\, \hat{c}(j,f(j)) \quad \text{for all } i \in I$$

where $\hat{p}_{ij}^0(f) = \delta_{ij}$. From (2.5), it follows that for any $f \in F$, $i, k \in I$ and $n \geq 1$ the total variation of the signed measure $\mu(A) = \Sigma_{j\in A}\hat{p}_{ij}^n(f) - \Sigma_{j\in A}\hat{p}_{kj}^n(f)$ is bounded by $4(1-\rho)^{[n/\nu]}$. Using this result and letting B any finite number such that $|\hat{c}(i,a)| \leq B$ for all i,a, it follows from (2.9) that, for any $0 < \beta < 1$ and all f $\in F$,

$$|\hat{V}_\beta(i,f^{(\infty)}) - \hat{V}_\beta(k,f^{(\infty)})| \leq 4B \sum_{n=0}^{\infty} (1-\rho)^{[n/\nu]} \leq \frac{4B\nu}{\rho} \text{ for all } i,k \in I.$$

Hence, by (2.8),

$$|\hat{V}_\beta(i) - \hat{V}_\beta(k)| \leq \frac{4B\nu}{\rho} \quad \text{for all } i,k \in I \text{ and all } 0 < \beta < 1.$$

Now, by using Lemma 2.3 and by making an obvious modification on the proof of Theorem 6.18 in [17], there exists a finite constant g and a bounded function v(i), $i \in I$ such that

$$(2.10) \qquad v(i) = \min_{a\in A(i)} \{\hat{c}(i,a) - g + \sum_{j\in I} \hat{p}_{ij}(a)v(j)\} \text{ for all } i \in I.$$

We shall now verify that $g^* = g$ and $v^*(i) = \tau v(i)$, $i \in I$ satisfy (2.6). To do this, observe that (2.10) can be equivalently written as

$$v(i) \geq \frac{c(i,a)}{\tau(i,a)} - g + \frac{\tau}{\tau(i,a)} \sum_{j\in I} p_{ij}(a)v(j) + (1 - \frac{\tau}{\tau(i,a)})\, v(i)$$
$$\text{for all } i \in I \text{ and } a \in A(i),$$

where for any $i \in I$ the equality holds for at least one $a \in A(i)$. Multiplying both sides of this inequality with $\tau(i,a) > 0$, we find

$$0 \geq c(i,a) - g\tau(i,a) + \tau \sum_{j\in I} p_{ij}(a)v(j) - \tau v(i)\,, \quad i \in I \text{ and } a \in A(i),$$

where for any $i \in I$ the equality sign holds for at least one $a\in A(i)$.

This proves that $g^* = g$ and $v^*(i) = \tau v(i)$, $i \in I$ satisfy the optimality equation (2.6). By Theorem 6.17 in [17], we have that the constant $g^*$ in (2.6) is uniquely determined and, by Lemma 3 in [12], we have that the function $v^*(i)$, $i \in I$ in (2.6) is uniquely determined up to an additive constant.

For any policy $\pi$, define for all $i \in I$ and $n \geq 1$,

$$V_n(i,\pi) = E_\pi[\sum_{k=0}^{n} c(X_k,a_k)|X_0=i] \text{ and } T_n(i,\pi) = E_\pi[\sum_{k=0}^{n} \tau(X_k,a_k)|X_0=i].$$

Define a policy $\pi^*$ to be *average cost optimal in the strong sense if*

$$(2.11) \qquad \limsup_{n\to\infty} \frac{V_n(i,\pi^*)}{T_n(i,\pi^*)} \leq \liminf_{n\to\infty} \frac{V_n(i,\pi)}{T_n(i,\pi)}$$

for all $i \in I$ and any policy $\pi$.

An examination of the proof Theorem 7.6 in[17] gives the following theorem.

THEOREM 2.2. Let $\{g^*, v^*(i), i\in I\}$ be any bounded solution to the optimality equation (2.6) and let $f_0 \in F$ be such that $f_0(i)$ minimizes the right side of (2.6) for all $i \in I$. Then

$$\liminf_{n\to\infty} \frac{V_n(i,\pi)}{T_n(i,\pi)} \geq g^* \text{ for all } i \in I \text{ and any policy } \pi$$

and

$$\lim_{n\to\infty} \frac{V_n(i,f_0^{(\infty)})}{T_n(i,f_0^{(\infty)})} = g^* \text{ for all } i \in I,$$

so the stationary policy $f_0^{(\infty)}$ is average cost optimal in the strong sense.

Observe that (2.11) implies $\limsup_{n\to\infty}\{V_n(i,\pi^*)/T_n(i,\pi^*) - V_n(i,\pi)/T_n(i,\pi)\} \leq 0$ for all $i \in I$ and any policy $\pi$. This latter average cost optimality criterion was considered in [8] where it was pointed out that this criterion is essentially stronger than both the lim sup and lim inf average cost criteria, cf. [2] - [4] in [8]. We note that in the literature the existence of an average cost optimal stationary policy is usually established under the lim sup average cost criterion.

Finally, letting $Z(t)$ be the total costs incurred up to time $t$ and using Theorem 7.5 in [17], we have under each of the conditions C1, C2 and C3 that, for any $f \in F$,

$$\lim_{t\to\infty} \frac{1}{t} E_{f^{(\infty)}}[Z(t)\,|\,X(0) = i] = \lim_{n\to\infty} \frac{V_n(i,f^{(\infty)})}{T_n(i,f^{(\infty)})} = g(f) \text{ for all } i \in I,$$

where g(f) is defined by

$$(2.12) \qquad g(f) = \Sigma_{i\in I}\, c(i,f(i))\pi_i(f)/\Sigma_{i\in I}\tau(i,f(i))\pi_i(f), \qquad f \in F,$$

with $\{\pi_j(f),\ j\in I\}$ is the unique stationary probability distribution of P(f).

## 3. THE VALUE-ITERATION METHOD.

In this section it is assumed that at least one of the conditions C1, C2 and C3 hold. We shall show that a bounded solution to the optimality equation (2.6) may be obtained by using value-iteration. In the proof of Theorem 2.1 we have found that any bounded solution $\{g, v(i),\ i\in I\}$ to the optimality equation for the transformed model gives a bounded solution $\{g^*=g,\ v^*(i)=\tau v(i),\ i\in I\}$ to the optimality equation (2.6). Hence, in view of the data-transformation given in section 2, it is no restriction to assume that $\tau(i,a) = 1$ for all i,a and P(f) is *aperiodic* for all $f \in F$.

Let $\{g^*,\ v^*(i),\ i\in I\}$ be any bounded set of numbers satisfying

$$(3.1) \qquad v^*(i) = \min_{a\in A(i)} \{c(i,a) - g^* + \sum_{j\in I} p_{ij}(a)v^*(j)\} \quad \text{for all } i \in I.$$

For any given bounded function $v_0(i)$, $i \in I$, define for n = 1,2,... the bounded function $v_n(i)$, $i \in I$ by the value-iteration equations

$$(3.2) \qquad v_n(i) = \min_{a\in A(i)} \{c(i,a) + \sum_{j\in I} p_{ij}(a)v_{n-1}(j)\} \text{ for } i \in I.$$

Observe that, by A3, the minimum in the right side of (3.2) is attained for all i. The asymptotic behaviour of the sequence $\{v_n(i)-ng^*,\ n\geq 1\}$ for $i \in I$ has been studied in [12] where the action sets A(i) were taken finite. This finiteness is in fact only used to verify relation (18) in [12], however, by invoking Lemma 32 on p. 178 in [18], it follows that the results in [12] also apply when for any $i \in I$ the set A(i) is a compact metric space such that both c(i,a) and $p_{ij}(a)$ for any $j \in I$ are continuous on A(i). Since the assumptions 1 - 5 in [12] are satisfied, we have for some constant c that

(3.3) $$\lim_{n\to\infty} \{v_n(i) - ng^*\} = v^*(i) + c \quad \text{for all } i \in I.$$

Hence, by choosing some state $i_0$ and defining $y_n=v_n(i_0)-v_{n-1}(i_0)$ and $w_n(i) = v_n(i)-v_n(i_0)$ for $i \in I$ and $n \geq 1$, it follows that the bounded numbers $\{y_n , w_n(i), i\in I\}$ converge as $n\to\infty$ to a bounded solution to (3.1). We further note that, by letting $f_n \in F$ be such that $f_n(i)$ minimizes the right side of (3.2) for all i and defining the average costs g(f) by (2.12), it follows by making minor modifications on standard arguments used in [10] and [16] that, for all $n \geq 1$,

$$\inf_{i\in I}\{v_n(i)-v_{n-1}(i)\} \leq g^* \leq g(f_n) \leq \sup_{i\in I}\{v_n(i)-v_{n-1}(i)\},$$

where $\inf_i\{v_n(i)-v_{n-1}(i)\}$ and $\sup_i\{v_n(i)-v_{n-1}(i)\}$ are non-decreasing and non-increasing respectively in $n \geq 1$.

Finally, consider the special case where condition C3 with $\nu=1$ holds. Let $B$ be the class of all bounded functions on I and define the mapping $T: B \to B$ by

$$Tu(i) = \min_{a\in A(i)} \{c(i,a) + \sum_{j\in I} p_{ij}(a)u(j)\}$$

and define $||u|| = \sup_i u(i) - \inf_i u(i)$ for $u \in B$. Then a repetition of the proof of Theorem 5 in [7] shows that, for some number $\rho > 0$, $|| Tu - Tw|| \leq (1-\rho) || u-w||$ for all $u,w \in B$ i.e. T is a contraction mapping. Next, using this result and the existence of a bounded solution to (3.1), it is readily verified that $|v_n(i)-ng^*-v^*(i)| \leq (1-\rho)^n || v_0-v^* ||$ for all $i \in I$ and $n \geq 1$, i.e. in this case the convergence in (3.3) is geometrically fast and uniform in i.

## 4. THE POLICY ITERATION METHOD.

Throughout this section it is assumed that condition C1 of section 1 holds and under this condition we shall study the convergence of the policy iteration method. Using ideas from a convergence proof given in [3] for a policy iteration approach to controlled Markov processes with a general state space, it will be shown that the average costs and the relative cost functions of the stationary policies generated by the policy iteration method converge to a bounded solution to the optimality equation (2.6).

As a byproduct we obtain an alternative proof of the existence of such a solution. Partial convergence results of this type were obtained in [4] under the restrictive additional assumption of no transient states under any $P(f)$, $f \in F$.

We first give some preliminary results. Let the state $s$ and the random variable $N$ be as in condition C1. For any $f \in F$, the stochastic matrix $P(f)$ has a unique stationary probability distribution $\{\pi_j(f), j \in I\}$ such that

$$(4.1) \qquad \pi_i(f) = \Sigma_{j \in I} p_{ji}(f) \pi_j(f) \quad \text{for all } i \in I.$$

Moreover, we have from Markov chain theory

$$(4.2) \qquad \pi_i(f) = \sum_{n=0}^{\infty} {}_s p_{si}^n(f) / E_{f^{(\infty)}} [N | X_0 = s] \quad \text{for all } i \in I$$

where ${}_s p_{ij}^0 = \delta_{ij}$ for all $i,j$ and

$$(4.3) \qquad {}_s p_{ij}^n(f) = P_{f^{(\infty)}}\{X_n = j,\ X_k \neq s \text{ for } 1 \le k \le n | X_0 = i\} \quad \text{for } i,j \in I \text{ and } n \ge 1.$$

Observe that

$$(4.4) \qquad E_{f^{(\infty)}}[N | X_0 = i] = 1 + \sum_{n=1}^{\infty} \sum_{j \in I} {}_s p_{ij}^n(f) \quad \text{for all } i \in I.$$

Further, for any $f \in F$, define the average costs $g(f)$ by (cf. (2.12))

$$(4.5) \qquad g(f) = \sum_{i \in I} c(i,f(i))\, \pi_i(f) / \sum_{i \in I} \tau(i,f(i))\, \pi_i(f),$$

and define the relative cost function $w_i(f)$ by

$$(4.6) \qquad w_i(f) = \sum_{n=0}^{\infty} \sum_{j \in I} \{c(j,f(j)) - g(f)\, \tau(j,f(j))\}\, {}_s p_{ij}^n(f), \ i \in I.$$

It is immediately verified from (4.2) and (4.4) - (4.6) that, for any $f \in F$, the function $w_i(f)$, $i \in I$ is bounded and has the property that

$$(4.7) \qquad w_s(f) = 0.$$

Consider now for fixed $f \in F$ the following system of linear equations in $\{g, v_i, i \in I\}$, (4.8) $$v_i = c(i, f(i)) - g\tau(i, f(i)) + \sum_{j\in I} p_{ij}(f(i))v_j \quad \text{for } i \in I.$$

We have the following known theorem (see [2] and [5]).

THEOREM 4.1. For any $f \in F$,

(a) The set of numbers $\{g = g(f),\ v_i = w_i(f),\ i\in I\}$ is a bounded solution to (4.8).

(b) For any bounded solution $\{g, v_i, i\in I\}$ to (4.8) holds $g = g(f)$.

(c) For any two bounded solutions $\{g, v_i\}$ and $\{g, u_i\}$ to (4.8) there is a constant c such that $v_i - u_i = c$ for all $i \in I$.

(d) For any $j \in I$, there is a unique bounded solution $\{g, v_i\}$ to (4.8) such that $v_j = 0$.

REMARK 4.1. To verify Theorem 4.1, it is not necessary to assume in A2 that $\inf_{i,a} \tau(i,a) > 0$ but it suffices to require that $\Sigma_i\ \tau(i,f(i))\ \pi_i(f) > 0$ for all $f \in F$.

By the assumptions A1-A2 and definition (4.5), we have

LEMMA 4.1. The set of numbers $\{g(f), f\in F\}$ is bounded.

Actually we shall only need that the numbers $\{g(f), f\in F\}$ are bounded from below. Now, we have established this result we shall make no further use of the assumption that $\inf_{i,a}\tau(i,a) > 0$.

For any $f \in F$ and any bounded solution $\{g(f), v_i(f), i\in I\}$ to (4.8), define

(4.9) $$T(i,a, v(f)) = c(i,a) - g(f)\tau(i,a) + \sum_{j\in I} p_{ij}(a)v_j(f) \quad \text{for } i \in I \text{ and } a \in A(i).$$

Observe that

(4.10) $$T(i, f(i), v(f)) = v_i(f) \quad \text{for all } i \in I \text{ and } f \in F.$$

The following lemma shows how the stationary policy $f^{(\infty)}$ can be improved to a stationary policy $h^{(\infty)}$ whose average costs is less than or equal to that of $f^{(\infty)}$.

LEMMA 4.2. Let $f \in F$ and let $\{g(f), v_i(f)\}$ be any bounded solution to (4.8). Suppose $h \in F$ is such that

$$(4.11) \qquad T(i, h(i), v(f)) \leq v_i(f) \quad \text{for all } i \in I.$$

Then $g(h) \leq g(f)$.

PROOF. The proof is standard. Multiply both sides of the inequality (4.11) with $\pi_i(f)$ and sum over $i \in I$. Next the desired result follows after an interchange of the order of summation which is justified by the boundedness of $v(f)$ and using the steady-state equation (4.1) for policy $h^{(\infty)}$.

We now formulate the policy-iteration method.

*Policy Iteration Method*

Step 0. Initialize with any $f_1 \in F$.

Step 1. Let $f^{(\infty)}$ be the current policy. Determine the unique bounded solution $\{g(f), w_i(f)\}$ to the system of linear equations (4.8) in which $v_s = 0$.

Step 2. Determine $f' \in F$ such that $T(i, f'(i), w(f)) = \min_{a \in A(i)} T(i,a,w(f))$ for all $i \in I$ where $f'(i)$ is chosen equal to $f(i)$ when this action minimizes $T(i,a, w(f))$ for $a \in A(i)$. Go to step 1.

Let $\{f_n^{(\infty)}, n \geq 1\}$ be the sequence of stationary policies generated by the policy iteration method. Observe that, by part (c) of Theorem 4.1, $f_{n+1}$ is independent of the particular choice of the bounded solution to (4.8) with $f = f_n$. By Lemma 4.2,

$$(4.12) \qquad g(f_{n+1}) \leq g(f_n) \quad \text{for all } n \geq 1.$$

We shall prove that the bounded numbers $\{g(f_n), w_i(f_n), i \in I\}$ converge as $n \to \infty$ to a bounded solution to the optimality equation (2.6). To do this, we shall use a modified semi-Markov decision model specified by the five objects $(\bar{I}, \bar{A}(i), \bar{p}_{ij}(a), \bar{c}(i,a), \bar{\tau}(i,a))$ where, for some artificial state $\infty$ and action $a_\infty$ (say),

$$\bar{I} = I \cup \{\infty\}, \ \bar{A}(i) = A(i) \text{ for } i \in I, \ \bar{A}(\infty) = \{a_\infty\},$$

$$\bar{c}(i,a) = c(i,a),\ \bar{\tau}(i,a) = \tau(i,a) \text{ for } i \in I \text{ and } a \in A(i),$$

$$\bar{c}(\infty,a_\infty) = \bar{\tau}(\infty,a_\infty) = 0,\ \bar{p}_{\infty s}(a_\infty) = 1,\ \bar{p}_{\infty j}(a_\infty) = 0 \text{ for } j \neq s,$$

$$\bar{p}_{ij}(a) = \begin{cases} p_{ij}(a) \text{ for } i,j \in I,\ a \in A(i),\ j \neq s \\ p_{sj}(a) \text{ for } i \in I,\ a \in A(i),\ j = \infty. \end{cases}$$

In fact this modified model is identical to the semi-Markov decision model considered except that before any transition to state s there first occurs a transition to state $\infty$ after which an instantaneous transition occurs to state s involving no costs. For the modified model, denote by $\bar{F}$ the class of all functions h which add to each state $i \in \bar{I}$ a single action $h(i) \in \bar{A}(i)$ and associate with any $h \in \bar{F}$ the stochastic matrix $\bar{P}(h) = (\bar{p}_{ij}(h(i)))$, $i,j \in \bar{I}$. Since $h(\infty) = a_\infty$ for all $h \in \bar{F}$, there is a one-to-one correspondence between F and $\bar{F}$. For any $f \in F$, denote by $\bar{f}$ the unique element in $\bar{F}$ such that $\bar{f}(i) = f(i)$ for all $i \in I$. It is immediate that there is a finite number B (say) such that under any stochastic matrix $\bar{P}(\bar{f})$, $f \in F$ the number of transitions it takes before the first return to state $\infty$ is bounded by B for any starting state $i \in \bar{I}$. Hence condition C1 with state s replaced by state $\infty$ also applies for the modified model. This result together with the fact that $\bar{A}(\infty)$ consists of a *single* action will play a crucial role in the convergence proof below. Further, for any $f \in F$, the stochastic matrix $\bar{P}(\bar{f})$ has a unique stationary probability distribution $\{\bar{\pi}_j(f),\ j \in \bar{I}\}$. Using the steady-state equation, we have for any $f \in F$ that $\bar{\pi}_s(f) = \bar{\pi}_\infty(f)$ and $\bar{\pi}_i(f) = \pi_i(f)/\{1+\pi_s(f)\}$ for all $i \in I$. Hence, letting

$$\bar{g}(f) = \sum_{i \in I} \bar{c}(i, \bar{f}(i))\bar{\pi}_i(f) / \sum_{i \in \bar{I}} \bar{\tau}(i, \bar{f}(i))\bar{\pi}_i(f) \qquad \text{for } f \in F$$

it follows that

$$(4.13) \qquad \bar{g}(f) = g(f) \text{ for all } f \in F.$$

Further, for any $f \in F$ define

$$\bar{w}_i(f) = \sum_{n=0}^{\infty} \sum_{j \in \bar{I}} \{\bar{c}(j, \bar{f}(j)) - \bar{g}(f)\,\bar{\tau}(j, \bar{f}(j))\}\,{}_{\infty}\bar{p}^{n}_{ij}(\bar{f}), \qquad i \in \bar{I}$$

where the definition of ${}_{\infty}\bar{p}^{n}_{ij}(\bar{f})$ is anologous to that of ${}_{s}p^{n}_{ij}(f)$ in (4.3). Then similarly as above, the bounded function $\bar{w}_i(f)$, $i \in \bar{I}$ has for any $f \in F$ the property

$$(4.14) \qquad \bar{w}_\infty(f) = 0$$

Since Theorem 4.1 also applies to the modified model, we have for any $f \in F$ that $\{g = \bar{g}(f),\ v_i = \bar{w}_i(f),\ i \in \bar{I}\}$ is the unique bounded solution to

$$(4.15) \qquad v_i = \bar{c}(i, \bar{f}(i)) - g\bar{\tau}(i, \bar{f}(i)) + \sum_{j\in\bar{I}} \bar{p}_{ij}(\bar{f}(i))v_j, \qquad i \in \bar{I},$$

having the property that $v_\infty = 0$. Further, using (4.13) - (4.15), it is immediately verified for any $f \in F$ that $\bar{w}_\infty(f) = \bar{w}_s(f)$ and that $\{g = g(f),\ v_i = \bar{w}_i(f),\ i \in I\}$ is a bounded solution to (4.8) having the property that $v_s = 0$. By the parts (a) and (d) of Theorem 4.1, it now follows that

$$(4.16) \qquad \bar{w}_i(f) = w_i(f) \quad \text{for all } i \in I \text{ and } f \in F.$$

Using the relations (4.14), it is now straightforward to verify that for any sequences $\{f_n^{(\infty)}, n\geq 1\}$ with $f_n \in F$ and $\{h_n^{(\infty)}, n\geq 1\}$ with $h_n \in \bar{F}$ generated by the policy-iteration method applied on the semi-Markov decision model considered and the modified model respectively, we have

$$(4.17) \qquad h_n = \bar{f}_n \quad \text{for all } n \geq 1 \text{ when } h_1 = \bar{f}_1.$$

The above relationships will be used to prove the convergence results for the policy-iteration method. Before doing this, we give the following lemma.

LEMMA 4.3. Let $\{u_n, n\geq 1\}$ be a bounded sequence of numbers such that for any $\varepsilon > 0$ there is an integer $N(\varepsilon)$ for which $u_{n+m} \leq u_n + \varepsilon$ for all $n,m \geq N(\varepsilon)$. Then the sequence $\{u_n\}$ is convergent.

PROOF. Let $u = \liminf_{n\to\infty} u_n$ and let $U = \limsup_{n\to\infty} u_n$. Choose $\varepsilon > 0$. Then, $U \leq u_n + \varepsilon$ for all $n \geq N(\varepsilon)$, so, $U \leq u + \varepsilon$ which proves the lemma since $\varepsilon$ was arbitrarily chosen.

We now prove the convergence results for the policy-iteration method.

THEOREM 4.2. Let $\{f_n^{(\infty)}, n\geq 1\}$ with $f_n \in F$ be any sequence of stationary policies generated by the policy-iteration method applied on the semi-Markov decision model considered. Then

$$(4.18)\qquad \lim_{n\to\infty} g(f_n) = \inf_{f\in F} g(f)$$

and, for some bounded function $w_i^*$, $i \in I$,

$$(4.19)\qquad \lim_{n\to\infty} w_i(f_n) = w_i^* \quad \text{for all } i \in I.$$

Moreover, letting $g^* = \inf_{f\in F} g(f)$, the bounded numbers $\{g^*, w_i^*, i\in I\}$ satisfy the optimality equation

$$(4.20)\qquad w_i^* = \min_{a\in A(i)} \{c(i,a) - g^*\tau(i,a) + \sum_{j\in I} p_{ij}(a)w_j^*\} \quad \text{for all } i \in I.$$

PROOF. Suppose that we have already verified (4.18) and (4.19). Using the construction of $f_n$ and the relations (4.8) and (4.9), we have for all $n \geq 2$

$$(4.21)\qquad w_i(f_n) = c(i, f_n(i))-g(f_n)\tau(i,f_n(i)) + \sum_{j\in I} p_{ij}(f_n(i))w_j(f_n),\ i \in I$$

and

$$(4.22)\qquad c(i, f_n(i))-g(f_{n-1})\tau(i, f_n(i)) + \Sigma p_{ij}(f_n(i))w_j(f_{n-1}) =$$

$$= \min_{a\in A(i)} \{c(i,a)-g(f_{n-1})\tau(i,a) + \sum_{j\in I} p_{ij}(a)w_j(f_{n-1})\}. \qquad i \in I.$$

Since I is denumerable and A(i) is a compact metric space for any $i \in I$, we can choose a $f^* \in F$ and an infinite sequence $\{n_k, k\geq 1\}$ such that

$$\lim_{k\to\infty} f_{n_k}(i) = f^*(i) \quad \text{for all } i \in I.$$

Now, taking $n = n_k$ in (4.21) and (4.22), letting $k \to \infty$ and using A3 together with the same arguments as in the proof of Lemma 2.3, we easily get the result (4.20) where $f^*(i)$ minimizes the right-side of (4.20) for all $i \in I$. It remains to prove (4.18) and (4.19). We shall first prove these relations under the assumption

(4.23) the action set A(s) consists of a single action.

Next, using the modified model, we shall verify that (4.18) and (4.19) also hold without the assumption (4.23). Now suppose that (4.23) holds. Fix $n \geq 1$. By (4.23), we have $f_{n+1}(s) = f_n(s)$ and so, by (4.9) and part (a) of Theorem 4.1,

$$T(s,f_{n+1}(s),w(f_n)) = c(s,f_n(s))-g(f_n)\tau(s,f_n(s))+\sum_{j\in I} p_{sj}(f_n(s))w_j(f_n) = w_s(f_n).$$

Hence, by (4.7),

$$(4.24) \qquad T(s,f_{n+1}(s),w(f_n)) = 0.$$

Put for abbreviation

$$a_n(i) = c(i,f_{n+1}(i)) - g(f_n)\tau(i,f_{n+1}(i)) \quad \text{for } i \in I.$$

Then, by (4.9),

$$(4.25) \qquad T(i,f_{n+1}(i),w(f_n)) = a_n(i)+\sum_{j\in I} p_{ij}(f_{n+1})w_j(f_n) \quad \text{for } i \in I.$$

By the construction of $f_{n+1}$ and (4.10),

$$(4.26) \qquad w_j(f_n) \geq T(j,f_{n+1}(j),w(f_n)) \quad \text{for all } j \in I.$$

Using (4.24) - (4.26) and (4.3), we have for any $i \in I$

$$T(i,f_{n+1}(i),w(f_n)) \geq a_n(i)+\sum_{j\neq s} p_{ij}(f_{n+1})T(j,f_{n+1}(j),w(f_n)) =$$

$$= a_n(i) + \sum_{j\in I} {}_sp^1_{ij}(f_{n+1})T(j,f_{n+1}(j),w(f_n)) =$$

$$= a_n(i) + \sum_{j\in I} {}_sp^1_{ij}(f_{n+1})a_n(j) + \sum_{j\in I} {}_sp^1_{ij}(f_{n+1}) \sum_{h\in I} p_{jh}(f_{n+1})w_h(f_n)$$

Continuing in this way, we find by induction on m that for any $m \geq 1$

$$T(i,f_{n+1}(i),w(f_n)) \geq \sum_{k=0}^{m} \sum_{j\in I} a_n(j)\; {}_sp^k_{ij}(f_{n+1}) +$$

$$+ \sum_{j\in I} {}_sp^m_{ij}(f_{n+1}) \sum_{h\in I} p_{jh}(f_{n+1})w_h(f_n), \quad i \in I.$$

We now observe that, by condition C1 and relation (4.4),

$$\lim_{m\to\infty} \sum_{j\in I} {}_s p_{ij}^m(f) = 0 \quad \text{for all } i \in I \text{ and } f \in F.$$

Using this result, (4.4) and the boundedness of the functions $a_n(i)$ and $w_i(f_n)$, $i \in I$, it now follows that

$$(4.27) \qquad T(i,f_{n+1}(i),w(f_n)) \geq \sum_{k=0}^{\infty} \sum_{j\in I} a_n(j)\ {}_s p_{ij}^k(f_{n+1}) \quad \text{for all } i \in I.$$

Putting $\Delta_n = g(f_n) - g(f_{n+1})$, it follows from (4.6) and (4.27) that, for any $i \in I$,

$$w_i(f_{n+1})-T(i,f_{n+1}(i),w(f_n)) \leq \Delta_n \sum_{k=0}^{\infty} \sum_{j\in I} \tau(j,f_{n+1}(j))\ {}_s p_{ij}^k(f_{n+1}),$$

where the various operations on the sums involved are justified by the absolute convergence of these sums. Next, using the boundedness of $\tau(i,a)$, relation (4.4) and condition C1, there is some finite number B such that

$$(4.28) \qquad w_i(f_{n+1})-T(i,f_{n+1}(i),w(f_n)) \leq \Delta_n B \quad \text{for all } i \in I \text{ and } n \geq 1.$$

Hence, by (4.26) and (4.28), $w_i(f_{n+1})-w_i(f_n) \leq \Delta_n B$ for all $i \in I$ and $n \geq 1$ which implies

$$(4.29) \qquad w_i(f_{n+m})-w_i(f_n) \leq \{g(f_n)-g(f_{n+m})\}B \quad \text{for all } i \in I \text{ and } n,m \geq 1.$$

Since the sequence $\{g(f_n), n\geq 1\}$ is bounded from below and non-increasing (see Lemma 4.1 and (4.12)), it follows that $\lim_{n\to\infty} g(f_n)$ exists and is finite. Next, using (4.29) and Lemma 4.3, we obtain (4.19) for some bounded function $w_i^*$, $i \in I$. To prove (4.18), observe that, by (4.26),

$$0 \leq w_i(f_n)-w_i(f_{n+1})+w_i(f_{n+1})-T(i,f_{n+1}(i),w(f_n)) \quad \text{for all } i \in I \text{ and } n \geq 1,$$

and so, by (4.19) and (4.28),

$$(4.30) \qquad \lim_{n\to\infty} \{w_i(f_n)-T(i,f_{n+1}(i),w(f_n))\} = 0 \quad \text{for all } i \in I.$$

Choose now $f \in F$. By the definition of $f_{n+1}$ and (4.9), we have for all $i \in I$ and $n \geq 1$,

$$c(i,f(i))-g(f_n)\tau(i,f(i))+ \sum_{j\in I} p_{ij}(f)w_j(f_n) \geq T(i,f_{n+1}(i),w(f_n))-w_i(f_n)+w_i(f_n).$$

Multiply both sides of this inequality with $\pi_i(f)$ and sum over $i \in I$. After an interchange of the order of summation justified by the boundedness of the functions involved and using (4.1), we get

$$\sum_{i \in I} \{c(i,f(i))-g(f_n)\tau(i,f(i))\}\pi_i(f) \geq \sum_{j \in I} \{T(i,f_{n+1}(i),w(f_n))-w_i(f_n)\}\pi_i(f).$$

Next, letting $n \to \infty$ and using the bounded convergence theorem and the relations (4.30) and (4.5), we find $g(f) \geq \lim_{n\to\infty} g(f_n)$ which implies (4.18) since $f \in F$ was arbitrarily chosen. We now have verified (4.18) and (4.19) under the assumption (4.23). Finally, using the modified model for which condition C1 with state $\infty$ in stead of state s applies and $\bar{A}(\infty)$ consists of a single action, and using the relations (4.13), (4.14), (4.16) and (4.17), the above proof shows that (4.18) and (4.19) also hold without the assumption (4.23). This completes the proof.

## REFERENCES.

1. ANTHONISSE, J.M. and TIJMS, H.C. (1975). Exponential convergence of products of stochastic matrices, Report BW 58/75, Mathematisch Centrum, Amsterdam (to appear in *J. Math. Anal. Appl.*).
2. DE LEVE, G., FEDERGRUEN, A. and TIJMS, H.C., (1976), A general Markov decision method,I: model and method, Report BW 61/76, Mathematisch Centrum, Amsterdam (to appear in *Adv. Appl. Prob.*).
3. ------------, (1977), *Generalized Markovian Decision Processes, Revisited,* Mathematical Centre Tract, Mathematisch Centrum, Amsterdam (in preparation).
4. DERMAN, C. (1966), Denumerable state Markovian decision processes-average cost criterion, *Ann. Math. Statist.* 37, 1545-1553.
5. DERMAN, C. and VEINOTT, A.,Jr. (1967), A solution to a countable system of equations arising in Markovian decision processes, *Ann. Math. Statist.* 38, 582-584.
6. DOOB, J.L. (1953), *Stochastic Processes,* Wiley, New York.
7. FEDERGRUEN, A., SCHWEITZER, P.J. and TIJMS, H.C. (1977), Contraction mappings underlying undiscounted Markov decision problems, Report BW 72/77, Mathematisch Centrum, Amsterdam.
8. FLYNN, J. (1976), Conditions for the equivalence of optimality criteria in dynamic programming (to appear in *Ann. Statist.*).
9. HAJNAL, J. (1958), Weak ergodicity in non homogeneous Markov chains, *Proc. Cambridge Philos. Soc.* 54, 233-246.
10. HASTINGS, N.A.J. (1971), Bounds on the gain of Markov decision processes, *Operations Res.* 10, 240-243.
11. HORDIJK, A. (1974), *Dynamic Programming and Potential Theory,* Mathematical Centre Tract No. 51, Mathematisch Centrum, Amsterdam.
12. HORDIJK, A., SCHWEITZER, P.J. and TIJMS, H.C. (1975), The asymptotic behaviour of the mimimal total expected cost for the denumerable state Markov decision model, *J. Appl. Prob.* 12, 298-305.
13. HORDIJK, A. and SLADKŶ, K. (1975), Sensitive optimality criteria in countable state dynamic programming, Report BW 48/75, Mathematisch Centrum, Amsterdam.
14. LIPPMAN, S.A. (1975), On dynamic programming with unbounded rewards, *Management Sci.* 21, 1225-1233.
15. MAITRA, A. (1968), Discounted dynamic programming on compact metric spaces, *Sankhya Ser. A.* 30, 211-216.
16. ODONI, A.R. (1969), On finding the maximal gain for Markov decision processes, *Operations Res.* 17, 857-860.
17. ROSS, S.M. (1970), *Applied Probability Models with Optimization Applications,* Holden-Day, Inc., San Francisco.

18. ROYDEN, H.L. (1968), *Real Analysis* (2nd.ed.), MacMillan, New York.

19. SCHWEITZER, P.J. (1971), Iterative solution of the functional equations of undiscounted Markov renewal programming, *J. Math. Anal. Appl.* 34, 495-501.

20. TIJMS, H.C. (1975), On dynamic programming with arbitrary state space, compact action space and the average return as criterion, Report BW 55/75, Mathematisch Centrum, Amsterdam.