stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 92/78 SEPTEMBER
(DEPARTMENT OF OPERATIONS REASEARCH)

A. FEDERGRUEN, A. HORDIJK & H.C. TIJMS

DENUMERABLE STATE SEMI-MARKOV DECISION PROCESSES WITH UNBOUNDED COSTS, AVERAGE COST CRITERION

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: 90C45

Denumerable state semi-Markov decision processes with unbounded costs, average cost criterion*)

by

A. Federgruen, A. Hordijk**) & H.C. Tijms***)

ABSTRACT

This paper considers an undiscounted semi-Markov decision model with a denumerable state space and compact metric action sets where the one-step expected costs and transition times are allowed to be unbounded. Under a condition which, roughly speaking, requires the existence of a finite set such that the supremum over all stationary policies of the expected time and the total expected absolute cost incurred until the first return to this set are finite for any starting state, we shall verify the existence of a finite solution to the average costs optimality equation and the existence of an average cost optimal stationary policy. These results considerably generalize results so far obtained in the literature.

KEY WORDS & PHRASES: *Semi-Markov decision processes, denumerable state space, unbounded one-step costs, average costs, optimality equation, optimal stationary policy*

---

*) This report will be submitted for publication elsewhere

**) A. Hordijk, Rijksuniversiteit, Leiden

***) H.C. Tijms, Vrije Universiteit, Amsterdam

1. *Introduction*

We are concerned with a dynamic system which at decision epochs beginning with epoch 0 is observed to be in one of the states of a *denumerable* state space I and subsequently is controlled by choosing an action. For any state $i \in I$, the set $A(i)$ denotes the set of pure actions available in state i. If at any decision epoch the system is in state i and action $a \in A(i)$ is taken, then, regardless of the history of the system, the following happens:

(i) an immediate cost $c(i,a)$ is incurred

(ii) the time until the next decision epoch and the state at the next decision epoch are random with joint probability distribution function $Q(.,.|i,a)$.

For any $i \in I$ and $a \in I$, let

$$p_{ij}(a) = Q(\infty,j|i,a) \text{ for } j \in I \text{ and } \tau(i,a) = \sum_{j \in I} \int_0^\infty tQ(dt,j|i,a).$$

i.e. $p_{ij}(a)$ denotes the probability that the next state will be j and $\tau(i,a)$ denotes the unconditional mean time until the next decision epoch when action a is taken in state i. Observe that $\Sigma_{j \in I} p_{ij}(a) = 1$ for all i,a. We make the following assumption.

ASSUMPTION 1.

(a) *For any* $i \in I$, *the set* $A(i)$ *is a compact metric set.*

(b) *For any* $i \in I$, *both* $c(i,a)$, $p_{ij}(a)$ *for any* $j \in I$ *and* $\tau(i,a)$ *are continuous on* $A(i)$.

(c) *There is a number* $\varepsilon > 0$ *such that* $\tau(i,a) \geq \varepsilon$ *for all* $i \in I$ *and* $a \in A(i)$.

We now introduce some familiar notions. For n = 0,1,.., denote by X_n and a_n the state and the action at the n^{th} decision epoch (the 0^{th} decision epoch is at epoch 0). A policy π for controlling the system is any measurable rule which for each n specifies which action to choose at the n^{th} decision epoch given the current state X_n and the sequence $(X_0,a_0,\ldots,X_{n-1},a_{n-1})$ of past states and actions where the actions chosen may be randomised. A policy π is called *memoryless* when the actions chosen are independent of the history of the system except for the present state.

Define $\mathcal{R}$ as the class of all stochastic matrices $P = (p_{ij})$, $i,j \in I$ such that for any $i \in I$ the elements of the i^{th} row of P can be represented by

$$(1.1) \quad p_{ij} = \int_{A(i)} p_{ij}(a)\, \pi_i(da) \text{ for all } j \in I$$

for some probability distribution $\pi_i\{.\}$ on $A(i)$. Then any memoryless policy π can be represented by some sequence $(P_1, P_2, ..)$ in $\mathcal{R}$ such that the i^{th} row of P_n gives the probability distribution of the state at the n^{th} decision epoch when the current state at the (n-1)st decision epoch is i and policy π is used. Define $F = X_{i \in I} A(i)$. Observe that, under assumption 1(a), F is a compact metric set in the product topology. For any $f \in F$, let $P(f)$ be the stochastic matrix whose (i,j)th element is $p_{ij}(f(i))$, $i,j \in I$ and for $n = 1,2,...$ denote by the stochastic matrix $P^n(f) = (p^n_{ij}(f))$ the n-fold matrix product of $P(f)$ with itself. A memoryless policy $\pi = (P_1, P_2, ...)$ is called *stationary* when $P_n = P$ for for all $n \geq 1$ and $P = P(f)$ for some $f \in F$. This policy which always prescribes to take the single action $f(i) \in A(i)$ whenever the system is in state i will be denoted by $f^{(\infty)}$. Observe that under the stationary policy $f^{(\infty)}$ the process $\{X_n, n \geq 0\}$ is a Markov chain with one-step transition probability matrix $P(f)$.

For $n = 0,1,...,$ denote by τ_n the time between the n^{th} and (n+1)st decision epoch. A policy π^* is said to be *(strongly) average cost optimal* when $\limsup_{n\to\infty} \phi_n(i,\pi^*)$ is less than or equal to $\limsup_{n\to\infty} \phi_n(i,\pi)$ $(\liminf_{n\to\infty} \phi_n(i,\pi))$ for any $i \in I$ and policy π where $\phi_n(i,\pi)$ is defined by

$$(1.2) \quad \phi_n(i,\pi) = \frac{E_\pi\{\Sigma_{k=0}^n c(X_k, a_k) \mid X_0 = i\}}{E_\pi\{\Sigma_{k=0}^n \tau_k \mid X_0 = i\}}, \; n = 0,1,...$$

with E_π is the expectation under policy π. We here assume that this quantity is well-defined for any $i \in I$ and policy π as is the case under the additional assumption 2(a) to be stated below.

It is well-known that an average cost optimal policy may not exist and even an example has been given in [7] in which an average cost policy exists but any average cost optimal policy is nonstationary. It is remarkable in this example, that besides uniformly bounded $c(i,a)$ and $\tau(i,a)$, any stochastic matrix $P \in \mathcal{R}$ is irreducible and positive recurrent.

In general we can only state that for fixed initial state we may restrict ourselves to the class of memoryless policies. More precisely, by a slight generalization of the proof of Theorem 2 in [2], we have the known result that for any fixed $i_0 \epsilon I$ and policy π_0 a memoryless policy π_M can be found such that for any $k \epsilon I$, Borel set $B \subseteq A(k)$ and $n \geq 0$,

$$(1.3) \quad Pr_{\pi_M}\{X_n = k,\ a_n \epsilon B | X_0 = i_0\} = Pr_{\pi_0}\{X_n = k,\ a_n \epsilon B | X_0 = i_0\}.$$

We also have as general result that if a finite constant g and a finite function v(i), $i \epsilon I$ exist satisfying the *average cost optimality equation*

$$(1.4) \quad v(i) = \min_{a \epsilon A(i)} \{c(i,a) - g\tau(i,a) + \sum_{j \epsilon I} p_{ij}(a) v(j)\} \quad \text{for all } i \epsilon I,$$

then under an additional condition on the function v(.) any stationary policy $f^{(\infty)}$ such that the pure action f(i) minimizes the right side of (1.4) for all $i \epsilon I$ is strongly average cost optimal.

We shall focus our attention on the existence of a finite solution to the average cost optimality equation and the existence of a strongly average cost optimal stationary policy. So far most of the existing literature has dealt with these questions both under the severe assumption of uniformly bounded functions c(i,a) and $\tau(i,a)$ and under very strong recurrence conditions on the stochastic matrices P(f), $f \epsilon F$, cf. [3]-[5], [9] and [13]. The recurrence condition in [4]-[5] required the unichainedness of the stochastic matrices P(f), $f \epsilon F$ and the existence of a finite set K such that the supremum over the stationary policies of the mean recurrence time to this set K is bounded in the starting state where in [3] and [13] the special case of K equal to a singleton was considered. However, the assumption of mean recurrence times that are bounded in the starting state is too strong for many applications as in inventory and queueing theory where also the one-step costs c(i,a) are usually unbounded. A treatment of the average cost model with unbounded c(i,a) has been given in [9], [17] and [19]. This paper also allows for unbounded one-step costs and exhibits the existence of a finite solution to the average cost optimality equation and the existence of a strongly average cost optimal stationary policy under a condition which, roughly speaking, requires the unichainedness of the stochastic matrices P(f), $f \epsilon F$ and the existence of a finite set K such that the supremum over the stationary policies of both the expected time and the total expected absolute costs incurred until the first return to this set K are finite for any starting state. These results considerably generalize on the one hand results in [4]-[5] by relaxing both the assumption of uniformly bounded $c(i,a)$ and $\tau(i,a)$ and the assumption that the above mean recurrence times are bounded in the starting state. On the other hand, they generalize results in [9] where a Liapunov condition was considered which is in fact the above condition with the set K equal to a singleton. Under different but related assumptions the papers [17] and [19] only deal with the existence of an average cost optimal stationary policy for the discrete-time Markov decision model.

In section 2 we will give the main body of our analysis by first establishing relationships between the original and the decision processes embedded on the finite set K. Next in section 3 we will verify both the average cost optimality equation and the existence of a strongly average cost optimal stationary policy.

2. *Analysis of embedded decision processes.*

We first need some notation. For any set $A \subset I$, define the random variable

$$N(A) = \inf \{n \geq 1 | X_n \in A\},$$

i.e. $N(A)$ denotes the number of transitions until the first return to the set A where $N(A) = \infty$ if $X_n \notin A$ for all $n \geq 1$. Also, for any $A \subset I$ and $f \in F$, define the taboo probability

$$(2.1) \qquad {}_A p^n_{ij}(f) = \Pr_{f^{(\infty)}} \{X_n = j, X_k \notin A \text{ for } 1 \leq k \leq n-1 | X_0 = i\},$$

$$i,j \in I \text{ and } n = 1,2,\ldots .$$

Observe that

$$(2.2) \qquad E_{f^{(\infty)}}\{N(A)|X_0 = i\} = 1 + \sum_{n=1}^{\infty} \sum_{j \notin A} {}_A p^n_{ij}(f).$$

We now introduce our main assumption.

ASSUMPTION 2.

(a) *There is a finite set* K *such that for any* $i \in I$ *the quantities* $u^*(i)$ *and* $y^*(i)$ *are finite where*

$$\sup_{f \in F} E_{f^{(\infty)}}\{\sum_{k=0}^{N(K)-1} \tau_k | X_0 = i\} = u^*(i) \text{ for all } i \in I \tag{2.3}$$

and

$$\sup_{f \in F} E_{f^{(\infty)}}\{\sum_{k=0}^{N(K)-1} |c(X_k, a_k)| \,|\, X_0 = i\} = y^*(i) \text{ for all } i \in I. \tag{2.4}$$

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

In words, assumption 2(a) requires the existence of a finite set K such that the supremum over all stationary policies of both the expected time and the total expected absolute cost incurred until the first return to the set K are finite for any starting state.

We shall now first verify as key result that under the assumptions 1-2 for any $f \in F$ a state $s_f \in K$ exists such that under policy $f^{(\infty)}$ the expected time and the total expected absolute cost incurred until the first return to the state s_f are bounded by $u^*(i) + c$ and $y^*(i) + c$ respectively for any starting state i for some constant c independent of $f \in F$. We shall need the following lemma.

LEMMA 2.1. *Let* A *be any subset of* I. *Then, for any* $i \in I$ *and* $f \in F$

$$E_{f^{(\infty)}}\{\sum_{k=0}^{N(A)-1} \tau(X_k, a_k) | X_0 = i\} = E_{f^{(\infty)}}\{\sum_{k=0}^{N(A)-1} \tau_k | X_0 = i\} = \tag{2.5}$$

$$= \tau(i, f(i)) + \sum_{n=1}^{\infty} \sum_{j \notin A}^{\infty} \tau(j, f(j))\, {}_A p_{ij}^n(f).$$

PROOF. Fix $i \in I$ and $f \in F$. For k=1,2,..., define the random variable

$$\delta_k = \begin{cases} 1 & \text{if } X_m \notin A \text{ for } 1 \le m \le k \\ 0 & \text{otherwise.} \end{cases}$$

Then, using the nonnegativity of τ_k and δ_k,

$$E\{\sum_{k=0}^{N(A)-1} \tau_k | X_0=i\} = E\{\tau_0 + \sum_{k=1}^{\infty} \tau_k\delta_k | X_0=i\} = \tau(i,f(i)) + \sum_{k=1}^{\infty} E\{\tau_k\delta_k | X_0=i\} =$$

$$= \tau(i,f(i)) + \sum_{k=1}^{\infty} \sum_{j\notin A} \tau(j,f(j)) \ {}_A p_{ij}^{n}(f).$$

By the same arguments, we have that the first expression in (2.5) equals the last one in (2.5) which verifies the relation (2.5).

By this lemma, we may replace τ_k by $\tau(X_k,a_k)$ in (2.3). This result will be essentially used in the analysis hereafter. It now follows from Lemma 2.1 and (2.2)-(2.3) that under the assumptions 1(c) and 2(a),

$$(2.6) \qquad E_{f^{(\infty)}}\{N(K) | X_0=i\} \le \frac{u^*(i)}{\varepsilon} \quad \text{for any } f\in F \text{ and } i\in I$$

REMARK. In [4]-[5] the uniformly boundedness of the functions $c(i,a)$ and $\tau(i,a)$, the unichainedness of the stochastic matrices $P(f)$, $f\in F$ and the so-called simultaneous Doeblin condition were assumed. This recurrence condition assumes that a finite set K, an integer $\nu \ge 1$ and a number $\rho > 0$ exist such that $\Sigma_{j\in K} p_{ij}^{\nu}(f) \ge \rho$ for all $f\in F$ and $i\in I$. Since this condition is equivalent to the condition requiring that, for some finite set K and constant B, $E_{f^{(\infty)}}\{N(K) | X_0=i\} \le B$ for all $f\in F$ and $i\in I$ (cf. [4] and [9]), we have that assumption 2 is satisfied with bounded functions u^* and y^* under the conditions considered in [4]-[5]. We note that in [6] several recurrence conditions were studied which are equivalent to the simultaneous Doeblin condition.

Under assumption 2, define for any $f\in F$

$$(2.7) \qquad q_{ij}(f) = \sum_{n=1}^{\infty} {}_K p_{ij}^{n}(f), \; i\in I, \; j\in K,$$

i.e. $q_{ij}(f)$ is the probability that at the first return to the set K the transition occurs into state j starting from state i and using policy $f^{(\infty)}$. Observe that, by (2.6), (2.8) $\sum_{j\in K} q_{ij}(f) = 1$ for all $i\in I$.

For any $f\in F$, define for $i\in I$ and $j\in K$ the (possibly infinite) quantity

(2.9) $\nu_{ij}(f)$ = expected number of returns to the set K until the first transition into state j occurs starting from state i and using policy $f^{(\infty)}$.

We now prove the following Theorem.

THEOREM 2.2. *Suppose that the assumptions* 1-2 *hold. Then*

(a) *For any* $f\in F$, *the finite stochastic matrix* $(q_{ij}(f))$, $i,j\in K$ *has no two disjoint closed sets.*

(b) *For any* $i\in I$ *and* $j\in K$, *the probability* $q_{ij}(f)$ *is continuous on* F.

(c) *There is a finite number* B *such that for any* $f\in F$ *a state* $s_f \in K$ *exists for which* $\nu_{is_f}(f) \le B$ *for all* $i\in I$.

PROOF. (a) Fix $f\in F$. Let $K_1 \subseteq K$ and $K_2 \subseteq K$ be any two non-empty sets that are closed under the stochastic matrix $Q(f) = (q_{ij}(f))$, $i,j\in K$. To prove that $K_1\cap K_2$ is not empty, define for r=1,2 the set

$$I_r = \{j\in I \mid p_{ij}^n(f) > 0 \text{ for some } i\in K_r \text{ and } n\ge 1\}.$$

It is immediate that both sets I_1 and I_2 are closed under P(f) and hence $I_1 \cap I_2 \ne \emptyset$. Choose now $t \in I_1 \cap I_2$. Since $t \in I_1$, it follows that

(2.10) $p_{st}^m(f) > 0$ for some $s\in K_1$ and $m\ge 1$.

We shall now verify that

(2.11) $p_{tu}^n(f) > 0$ for some $u\in K_2$ and $n\ge 1$.

To prove this, assume the contrary. Then, by $\Pr_{f^{(\infty)}}\{N(K)<\infty \mid X_0=t\} = 1$, we have $p_{tv}^k(f) > 0$ for some $v\in K\backslash K_2$ and $k\ge 1$. Since $t\in I_2$, it follows that $p_{it}^h(f) > 0$ for some $i\in K_2$ and $h\ge 1$ and so $p_{iv}^{h+k}(f)>0$. This implies that $v\in K\backslash K_2$ can be reached from state $i\in K_2$ under Q(f) contradicting that K_2 is closed under Q(f).

Hence (2.11) holds. By (2.10) and (2.11) $p_{su}^{m+n}(f) > 0$. This implies that $u \in K_2$ can be reached from state $s \in K_1$ under Q(f). Since K_1 is closed under Q(f), state u also belongs to K_1 so that $K_1 \cap K_2 \neq \emptyset$ as was to be verified.

(b) By assumption 1, we have that F is a compact metric set on which $p_{ij}(f)$ is continuous for any $i,j \in I$. Using this fact and the relation

$$_K p_{ij}^n(f) = \sum_{h \notin K} p_{ih}(f)\ _K p_{hj}^{n-1}(f) \quad \text{for } n = 2,3,\ldots$$

it follows by induction that $_K p_{ij}^n(f)$ is continuous on F for any $n \geq 1$ and $i,j \in I$. Hence $q_{ij}(f)$ is continuous on F if the sum (2.7) converges uniformly on F. To prove this, fix $s \in I$ and observe that, by (2.6),

$$(2.12) \qquad \sum_{n=0}^{\infty} \Pr_{f^{(\infty)}}\{N(K)>n \mid X_0=s\} \leq \frac{u^*(s)}{\varepsilon} \text{ for all } f \in F.$$

Choose now $0 < \delta < 1$. Then there is an integer M such that

$$(2.13) \qquad \Pr_{f^{(\infty)}}\{N(K) > M \mid X_0=s\} \leq \delta \quad \text{for all } f \in F.$$

To prove this, assume the contrary. Using the fact that $\Pr_{f^{(\infty)}}\{N(K)>n \mid X_0=s\}$ is non-increasing in n, we then get a contradiction with (2.12). Now, by (2.13) we have for any $j \in K$

$$\sum_{n=M+1}^{\infty} {}_K p_{sj}^n(f) \leq \Pr_{f^{(\infty)}}\{N(K)>M \mid X_0=s\} \leq \delta \quad \text{for all } f \in F$$

which proves the desired result since $\delta>0$ was chosen arbitrarily.

(c) By the finiteness of K and the assertions (a)-(b) of the Theorem, this assertion is an immediate consequence of Theorem 2.6 in [5] or Theorem 4 in [6].

The following theorem will play a crucial role in the analysis in the next section.

THEOREM 2.3. *Suppose that the assumptions* 1-2 *hold. Then there is a finite number* c *such that for any* $f \in F$ *a state* $s_f \in K$ *exists for which*

$$(2.14)\quad E_{f^{(\infty)}}\{\sum_{k=0}^{N(\{s_f\})-1} \tau(X_k,a_k)|X_0=i\} \le u^*(i)+c \text{ for all } i\in I$$

and

$$(2.15)\quad E_{f^{(\infty)}}\{\sum_{k=0}^{N(\{s_f\})-1} |c(X_k,a_k)||X_0=i\} \le y^*(i)+c \text{ for all } i\in I.$$

PROOF. By Theorem 2.2. we can choose a finite number B and for any $f\in F$ a state $s_f\in K$ such that

$$(2.16)\quad \nu_{is_f}(f) \le B \text{ for all } i\in I \text{ and } f\in F.$$

We shall now verify (2.14). The proof of (2.15) is identical. Fix now $f\in F$. We introduce the following notation. For any $i\in I$ and $j\in K$, define $\hat{q}^1_{ij}(f) = q_{ij}(f)$ and, for $n = 2,3,\ldots$, let

$$\hat{q}^n_{ij}(f) = \sum_{\substack{k\in K\\ k\neq s_f}} q_{ik}(f)\hat{q}^{n-1}_{kj}(f) \text{ for } i\in I \text{ and } j\in K.$$

Observe that $\hat{q}^n_{ij}(f)$ is the probability that during the first n-1 returns to the set K no transition occurs into state s_f and that at the n^{th} return to the set K a transition occurs into state j starting from state i and using policy $f^{(\infty)}$. We have

$$(2.17)\quad \nu_{is_f} = 1 + \sum_{n=1}^{\infty} \sum_{\substack{j\in K\\ j\neq s_f}} \hat{q}^n_{ij}(f) \text{ for all } i\in I.$$

Define $\nu_0=0$ and, for $n\ge1$, $\nu_n=\inf\{m>\nu_{n-1}|X_m\in K\}$. Also, define $\delta_0=1$ and, for any $k\ge1$, $\delta_k=1$ if $X_m\neq s_f$ for $1\le m\le k$ and $\delta_k=0$ otherwise. Denote by T(i,f) the first expression in (2.5) with A=K. Then using the first equality in (2.5) and (2.3), we find

$$E_{f^{(\infty)}}\{\sum_{k=0}^{N(\{s_f\})-1} \tau(X_k,a_k)|X_0=i\} = E_{f^{(\infty)}}\{\sum_{k=0}^{\infty} \delta_k \tau(X_k,a_k)|X_0=i\} =$$

$$= E_{f^{(\infty)}}\{\sum_{n=1}^{\infty} \sum_{k=\nu_{n-1}}^{\nu_n -1} \delta_k \tau(X_k,a_k)|X_0=i\} =$$

$$= T(i,f) + \sum_{n=2}^{\infty} E_{f^{(\infty)}}\{\sum_{k=\nu_{n-1}}^{\nu_n -1} \delta_k \tau(X_k,a_k)|X_0=i\} =$$

$$= T(i,f) + \sum_{n=2}^{\infty} \sum_{j\neq s_f} \hat{q}_{ij}^{n-1}(f)\ T(j,f) \leq$$

$$\leq u^*(i) + \max_{j\in K} u^*(j) \sum_{n=2}^{\infty} \sum_{j\neq s_f} \hat{q}_{ij}^{n-1}(f) \quad \text{for all } i\in I.$$

Invoking (2.16)-(2.17), we now get the desired result.

We now give some known results in positive dynamic programming (e.g. cf. [1], [9] and [16]). Since a directly accessible reference seems not be available, we include for completeness a simple proof.

LEMMA 2.4. *Consider the positive dynamic program* $(S,D(s),q(t|s,a), r(s,a))$ *where the state space* S *is denumerable, the action set* $D(s)$ *is a compact metric set for any* $s\in S$ *and the immediate return* $r(s,a)$ *is non-negative for all* $s\in S$ *and* $a\in D(s)$. *Also assume that for any* $s\in S$ *both* $r(s,a)$ *and the one-step transition probability* $q(t|s,a)$ *for any* $t\in S$ *are continuous on* $D(s)$. *For any policy* π, *define* $V(s,\pi) = E_\pi\{\Sigma_{n=0}^{\infty} r(X_n,a_n)|X_0=s\}$, $s\in S$ *where* X_n *and* a_n *denote the state and the action at the* n^{th} *decision epoch. Let* $V(s) = \sup_\pi V(s,\pi)$, $s\in S$. *Then*

$$\sup_{f\in F} V(s,f^{(\infty)}) = V(s) \text{ for all } s\in S \tag{2.18}$$

and

$$V(s) = \sup_{a\in D(s)} \{r(s,a) + \sum_{t\in S} V(t)q(t|s,a)\} \text{ for all } s\in S. \tag{2.19}$$

PROOF. We need some notation. For any integer $M\geq 1$, let $r^M(s,a) = \min(r(s,a),M)$ for all s,a. For any $0<\alpha<1$, $s\in S$ and policy π, define

$$V_\alpha(s,\pi) = E_\pi\{\sum_{n=0}^{\infty}\alpha^n r(X_n,a_n)|X_0=s\} \text{ and}$$

$$V_\alpha^M(s,\pi) = E_\pi\{\sum_{n=0}^{\infty}\alpha^n r^M(X_n,a_n)|X_0=s\}.$$

Using the non-negativity of $r(s,a)$ we have by the monotone convergence theorem

$$(2.20)\qquad \lim_{M\to\infty} V_\alpha^M(s,\pi) = V_\alpha(s,\pi) \text{ for any } 0<\alpha<1,\ s\in S \text{ and policy } \pi,$$

and, by a Tauberian theorem,

$$(2.21)\qquad \lim_{\alpha\to 1} V_\alpha(s,\pi) = V(s,\pi) \text{ for any } s\in S \text{ and policy } \pi.$$

Letting $V_\alpha^M(s) = \sup_\pi V_\alpha^M(s,\pi)$, $s\in I$, it is well-known from discounted dynamic programming (e.g. cf. [9] and [11]) that for any $0<\alpha<1$ and $M\geq 1$

$$(2.22)\qquad V_\alpha^M(s) = \max_{a\in D(s)}\{r(s,a)+\alpha\sum_{t\in S}V_\alpha^M(t)q(t|s,a)\} \text{ for all } s\in S$$

and

$$(2.23)\qquad \sup_{f\in F} V_\alpha^M(s,f^{(\infty)}) = \sup_\pi V_\alpha^M(s,\pi) \text{ for all } s\in S.$$

Using the fact that $\lim_{n\to\infty}\sup_x g_n(x) = \sup_x \lim_{n\to\infty} g_n(x)$ for any non-decreasing sequence of functions $\{g_n\}$, we obtain from (2.20) and (2.23) that $\sup_{f\in F}V_\alpha(s,f^{(\infty)}) = \sup_\pi V_\alpha(s,\pi)$ for all $s\in S$ and $0<\alpha<1$. Next, by letting $\alpha\to 1$ in this relation and using (2.21) we get (2.18). The optimality equation (2.19) follows by the same reasoning from (2.20)-(2.22) by first letting $M\to\infty$ and next letting $\alpha\to 1$.

We can now prove the following Theorem.

THEOREM 2.5. *Suppose that the assumptions* 1 *and* 2(a) *hold. Then*

(2.24) $$u^*(i) = \sup_{a\in A(i)} \{\tau(i,a) + \sum_{j\notin K} p_{ij}(a)u^*(j)\} \text{ for all } i\in I$$

(2.25) $$y^*(i) = \sup_{a\in A(i)} \{|c(i,a)| + \sum_{j\notin K} p_{ij}(a)y^*(j)\} \text{ for all } i\in I.$$

PROOF. To get (2.24), use the first equality in (2.5) and apply Lemma 2.3. with, for some artificial state ∞ and action a_∞,

$$S = I \cup \{\infty\},\ D(s) = A(s) \text{ for } s\in I,\ D(\infty) = \{a_\infty\},$$

$$r(s,a) = \tau(s,a) \text{ for } s\in I \text{ and } a\in D(s),\ r(\infty,a_\infty) = 0$$

$$q(t|s,a) = \begin{cases} p_{st}(a) & \text{for } s\in I,\ a\in D(s),\ t\in I\backslash K \\ 0 & \text{for } s\in I,\ a\in D(s),\ t\in K, \\ \sum_{t\in K} p_{st}(a) & \text{for } s\in I,\ a\in D(s),\ t=\infty \end{cases}$$

$$q(t|\infty,a_\infty) = \begin{cases} 1 & \text{for } t = \infty \\ 0 & \text{otherwise.} \end{cases}$$

In the same way we get (2.25) by taking $r(s,a) = |c(s,a)|$ for $s\in I$ and $a\in D(s)$.

By this theorem, we have that assumption 2(a) is equivalent to the condition requiring the existence of a finite set K and a finite non-negative function $y(i)$, $i\in I$ such that

(2.26) $$|c(i,a)| + \tau(i,a) + \sum_{j\notin K} p_{ij}(a)y(j) \le y(i) \text{ for all } i\in I \text{ and } a\in A(i).$$

The condition (2.26) with K equal to a singleton was first studied in [9] where this condition was called a Liapunov condition, cf. also [8] and [10] for further investigations on Liapunov conditions.

As a consequence of Theorem 2.5., we have that

(2.27) $$\sum_{j\notin K} p_{ij}(a)u^*(j) \le u^*(i) \text{ and } \sum_{j\notin K} p_{ij}(a)y^*(j) \le y^*(i) \text{ for all } i\in I \text{ and } a\in A(i).$$

For any stochastic matric $P = (p_{ij})$, $i,j\in I$, define the substochastic matrix $\hat{P} = (\hat{p}_{ij})$, $i,j\in I$ by

$$(2.28)\qquad \hat{p}_{ij} = \begin{cases} p_{ij} & \text{for } i\in I,\ j\notin K \\ 0 & \text{for } i\in I,\ j\in K \end{cases}$$

Then, by (2.27), we have for any $P \in R$ that

$$(2.29)\qquad \hat{P}u^*(i) \le u^*(i) \text{ and } \hat{P}y^*(i) \le y^*(i) \text{ for all } i\in I$$

where $Rx(i) = \Sigma_{j\in I} r_{ij}x(j)$ for any matrix $R = (r_{ij})$, $i,j\in I$ and function $x(.)$ on I. We conclude this section with the following Lemma.

LEMMA 2.6. *Suppose that the assumptions* 1 *and* 2(a) *hold. Then, for any sequence* $(P_1,P_2,\dots)$ *of stochastic matrices in* R,

$$(2.30)\qquad P_1\dots P_n u^*(i) \le u^*(i) + \max_{j\in K} u^*(j) + \sum_{k=1}^{n-1}\sum_{h\in K}\hat{P}_{k+1}\dots\hat{P}_n u^*(h) \le$$

$$\le u^*(i) + n\max_{j\in K} u^*(j) \text{ for all } n\ge 1 \text{ and } i\in I.$$

The same inequalities apply when u^* *is replaced by* y^*.

PROOF. By a last exit decomposition, we have for any $n\ge 1$, $i\in I$ and $j\notin K$,

$$(P_1\dots P_n)_{ij} = (\hat{P}_1\dots\hat{P}_n)_{ij} + \sum_{k=1}^{n-1}\sum_{h\in K}(P_1..P_k)_{ih}(\hat{P}_{k+1}..\hat{P}_n)_{hj}$$

By this relation and a repeated application of (2.29), we get (2.30).

We can now verify that under the assumptions 1 and 2(a) definition (1.2) makes sense. By Theorem 2.5., we have $|c(i,a)| \le y^*(i)$ for all $i\in I$ and $a\in A(i)$ and so, using assumption 1(c) and Lemma 2.6., it follows that for all $n\ge 1$ and $i\in I$ the quantity $\phi_n(i,\pi)$ is well-defined for any memoryless policy π and consequently, by (1.3), for any policy π.

3. *The average cost optimality equation.*

To derive the average cost optimality equation, we first analyse a discounted cost function, cf. [13] and [18]. For any $\beta>0$ and policy π, let

$$(3.1)\qquad V_\beta(i,\pi) = E_\pi\{\sum_{n=0}^{\infty} e^{-\beta\Sigma_{k=0}^{n-1}\tau(X_k,a_k)} c(X_n,a_n)|X_0=i\} \text{ for } i\in I.$$

We shall first verify that this quantity is well-defined under the assumptions 1 and 2(a). By (1.3) it suffices to verify this for the memoryless policies π. Choose any memoryless policy π and let π be represented by the sequence $(P_1,P_2,\ldots)$ of stochastic matrices in $\mathcal{R}$. By assumption 1(c) and (2.25), we have $\tau(i,a) \geq \varepsilon$ and $|c(i,a)| \leq y^*(i)$ for all $i\in I$ and $a\in A(i)$. Using Theorem 2.6., we now find

$$(3.2)\qquad E_\pi\{\sum_{n=0}^{\infty} |e^{-\beta\Sigma_{k=0}^{n-1}\tau(X_k,a_k)} c(X_k,a_k)||X_0=i\} \leq$$

$$\leq E_\pi\{\sum_{n=0}^{\infty} e^{-\beta n\varepsilon} y^*(X_n)|X_0=i\} = y^*(i) + \sum_{n=1}^{\infty} e^{-\beta n\varepsilon} P_1\ldots P_n y^*(i) \leq$$

$$\leq 2y^*(i) + (1-e^{-\beta\varepsilon})^{-2}\max_{j\in K} y^*(j) \quad \text{for all } i\in I.$$

This proves that $V_\beta(i,\pi)$ is well-defined. For any $\beta>0$, define

$$V_\beta(i) = \inf_\pi V_\beta(i,\pi) \text{ for } i\in I.$$

Then, since for any $\delta>0$ we can find a policy π_δ such that $V_\beta(i,\pi_\delta)-\delta \leq V_\beta(i) \leq V_\beta(i,\pi_\delta)$ for *all* $i\in I$, we have by (3.2) that for any $\beta>0$

$$(3.3)\qquad |V_\beta(i)| \leq 2y^*(i) + (1-e^{-\beta\varepsilon})^{-2} \max_{j\in K} y^*(j) \text{ for all } i\in I.$$

We now make the following assumption.

ASSUMPTION 3. *For any* $i\in I$, *both* $\Sigma_{j\in I}p_{ij}(a)u^*(j)$ *and* $\Sigma_{j\in I}p_{ij}(a)y^*(j)$ *are continuous on* $A(i)$.

By assumption 1 and a well-known convergence result, $\lim_{n\to\infty} \Sigma_{j\in A} p_{ij}(a_n) = \Sigma_{j\in A} p_{ij}(a)$ for any set $A\subseteq I$ if $\lim_{n\to\infty} a_n=a$. Hence, by (3.3) and the convergence theorem on p. 232 in [15], it follows that under the additional assumption 3 the function $\Sigma_{j\in I} p_{ij}(a)V_\beta(j)$ is continuous on A(i) for any $i\in I$. Using this result, a minor modification of the proof of Theorem 6.1 in [14] shows that for any $\beta>0$,

$$(3.4)\qquad V_\beta(i) = \min_{a\in A(i)} \{c(i,a) + e^{-\beta\tau(i,a)} \sum_{j\in I} p_{ij}(a)V_\beta(j)\} \text{ for all } i\in I.$$

Let $f_\beta^{(\infty)}$ be any stationary policy such that the action $f_\beta(i)$ minimizes the right side of (3.4) for all $i\in I$. Then

$$(3.5)\qquad V_\beta(i,f_\beta^{(\infty)}) = V_\beta(i) \text{ for all } i\in I.$$

To prove this, iterate $V_\beta(i) = c(i,f_\beta(i)) + e^{-\beta\tau(i,f_\beta(i))} \Sigma_{j\in I} p_{ij}(f_\beta)V_\beta(j)$, $i\in I$. This gives

$$(3.6)\qquad V_\beta(i) = E_{f_\beta^{(\infty)}}\{\sum_{n=0}^{m-1} e^{-\beta\Sigma_{k=0}^{n-1}\tau(X_k,a_k)} c(X_n,a_n)|X_0=i\} + $$
$$+ E_{f_\beta^{(\infty)}}\{e^{-\beta\Sigma_{k=0}^{m-1}\tau(X_k,a_k)} V_\beta(X_m)|X_0=i\} \quad \text{for all } m\geq 1 \text{ and } i\in I.$$

Using assumption 1(c), the inequality (3.3) and Lemma 2.6 , we find that for some constant c_β the second term in the right side of (3.6) is bounded by

$$2e^{-\beta m\varepsilon}\{P^m(f)y^*(i) + c_\beta\} \leq 2e^{-\beta m\varepsilon}\{y^*(i)+m \max_{j\in K} y^*(j)+c_\beta\}$$

for all $m\geq 1$ and $i\in I$.

Hence, by letting $m\to\infty$ in (3.6), we find (3.5). We now prove

LEMMA 3.1. *Suppose that the assumptions* 1-3 *hold. Then there are finite numbers* β^* , $\gamma > 0$ *such that for any* $f\in F$ *a state* $s_f\in K$ *exists for which*

$$(3.7)\qquad |\beta V_\beta(s_f,f^{(\infty)})| \leq \gamma \quad \textit{for all } 0 < \beta < \beta^*$$

and, for any $i\in I$,

(3.8) $|V_\beta(i,f^{(\infty)}) - V_\beta(s_f,f^{(\infty)})| \leq \gamma(u^*(i)+y^*(i))$ *for all* $0 < \beta < \beta^*$.

PROOF. By Theorem 2.3., we can choose a constant c and for any $f\in F$ a state $s_f\in K$ such that (2.14)-(2.15) hold. Fix now $\beta>0$ and $f\in F$. We have

$$(3.9)\quad V_\beta(i,f^{(\infty)}) = E_{f^{(\infty)}}\{\sum_{n=0}^{N(\{s_f\})-1} e^{-\beta\Sigma_{k=0}^{n-1}\tau(X_k,a_k)} c(X_n,a_n)|X_0=i\} +$$

$$+ E_{f^{(\infty)}}\{e^{-\beta\Sigma_{k=0}^{N(\{s_f\})-1}\tau(X_k,a_k)}|X_0=i\}V_\beta(s_f,f^{(\infty)}) \text{ for all } i\in I.$$

Taking $i=s_f$ in (3.9) and using (2.15) and assumption 1(c), we derive from (3.9) that

$$(3.10)\quad |V_\beta(s_f,f^{(\infty)})| \leq (y^*(s_f)+c)/(1-e^{-\beta\varepsilon}).$$

From (3.9), (2.14)-(2.15) and the inequality $1-e^{-x} \leq x$ for $x\geq 0$, we easily derive

$$(3.11)\quad |V_\beta(i,f^{(\infty)})-V_\beta(s_f,f^{(\infty)})| \leq y^*(i)+c+(u^*(i)+c)|\beta V_\beta(s_f,f^{(\infty)})| \text{for all } i\in I.$$

Since $\beta(1-e^{-\beta\varepsilon})^{-1}\to\varepsilon$ as $\beta\to 0$ and the set K is finite, we get the Lemma from (3.10)-(3.11).

We are now in a position to verify the average cost optimality equation.

THEOREM 3.2. *Suppose that the assumptions* 1-3 *hold. Then there is a constant* g *and a function* $v(i)$, $i\in I$ *with* $\sup_{i\in I}|v(i)|/\{u^*(i)+y^*(i)\} < \infty$ *such that*

$$(3.12)\quad v(i) = \min_{a\in A(i)} \{c(i,a)-g\tau(i,a) + \sum_{j\in I} p_{ij}(a)v(j)\} \text{ for all } i\in I.$$

PROOF. Following [13] and [18], fix some state $s\in I$. By (3.5) and Lemma 3.1, we can choose finite numbers β^*, $c>0$ such that

$$(3.13)\quad |\beta V_\beta(s)| \leq c \quad \text{for all } 0 < \beta < \beta^*$$

and for any $i\in I$,

$$(3.14)\quad |V_\beta(i)-V_\beta(s)| \leq c(u^*(i)+y^*(i)) \text{ for all } 0 < \beta < \beta^*.$$

For any $\beta>0$, let $f_\beta \in F$ be such that $f_\beta(i)$ minimizes the right side of (3.4) for all $i\in I$. Now, using (3.13)-(3.14), assumption 1(a) and the diagonalization method, we can find a sequence $\{\beta_k\}$ with $\beta_k \to 0$ as $k\to\infty$, a constant g, a function v(i), $i\in I$ and an action $a(i)\in A(i)$ for any $i\in I$ such that

$$(3.15)\qquad \lim_{k\to\infty} \beta_k V_{\beta_k}(s)=g,\ \lim_{k\to\infty} V_{\beta_k}(i)-V_{\beta_k}(s) = v(i) \text{ and } \lim_{k\to\infty} f_{\beta_k}(i) = a(i)$$

for all $i\in I$.

Observe that, by (3.14)-(3.15), $|v(i)|\le c(u^*(i)+y^*(i))$ for all $i\in I$. Now, subtracting $V_{\beta_k}(s)$ from both sides of (3.4) with $\beta=\beta_k$, letting $k\to\infty$, using the assumptions 1 and 3, relation (3.14) and the convergence theorem on p. 232 in [15], we obtain (3.12) in a standard way.

The assumptions 1-3 are satisfied in the example in [7] for which any average cost optimal policy is nonstationary. Hence an additional assumption is required to guarantee that a stationary policy $f^{(\infty)}$ such that the action f(i) minimizes the right side of (3.12) for all $i\in I$ is average cost optimal, cf. also [12]. We now state the following condition.

ASSUMPTION 4. *For any* $f\in F$, $\lim_{n\to\infty}\hat{P}^n(f)u^*(i)=\lim_{n\to\infty}\hat{P}^n(f)y^*(i) = 0$ *for all* $i\in I$ *where* $\hat{P}^n(f)$ *denotes the n-fold matrix product of the substochastic matrix* $\hat{P}(f)$ *defined by* (2.28) *with itself.*

We can now prove the following lemma.

LEMMA 3.3. *Suppose that the assumptions* 1, 2(a), 3 *and* 4 *hold. Then, for any sequence* $(P_1,P_2,\dots)$ *of stochastic matrices in* R,

$$(3.16)\qquad \lim_{n\to\infty} \frac{1}{n} P_1 \dots P_n(u^*+y^*)(i) = 0 \text{ for all } i\in I.$$

PROOF. Following the proof of Lemma 5.7. in [9], define $x_0(i)=u^*(i)+y^*(i)$, $i\in I$, and for n=1,2,..., define $x_n(i)$ recursively by

$$(3.17)\qquad x_n(i) = \sup_{a\in A(i)} \sum_{j\notin K} p_{ij}(a)x_{n-1}(j), \quad i\in I.$$

Using the assumptions 1 and 3, the convergence theorem on p. 232 in [15] and (2.27), we find by induction that we may replace sup by max in (3.17) and that $x_n(i) \le x_{n-1}(i)$ for all $n \ge 1$ and $i \in I$. Hence

(3.18) $\lim_{n\to\infty} x_n(i) = x(i)$ (say) exists for all $i \in I$

and for any $i \in I$ and $n \ge 1$ we can choose an action $a_n(i)$ which maximizes the right side of (3.17). By assumption 1(a), we can find a sequence $\{n_k\}$ of integers with $n_k \to \infty$ as $k \to \infty$ such that $a_{n_k}(i) \to a(i)$ as $k \to \infty$ for some $a(i) \in A(i)$ for any $i \in I$. Let $f_0 \in F$ be such that $f_0(i) = a(i)$ for all i. Then, using assumption 3 and the convergence theorem on p. 231 in [15], we get from (3.17)-(3.18) that $x(i) = \hat{P}(f_0)x(i)$ for all i. Hence, using $x \le x_0$, we have $0 \le x = \hat{P}^n(f_0)x \le \hat{P}^n(f_0)(u^* + y^*)$ for all $n \ge 1$ so that, by assumption 4,

(3.19) $x(i) = 0$ for all $i \in I$.

Now, let $(P_1, P_2, \ldots)$ be any sequence of stochastic matrices in $\mathcal{R}$. By (3.17), $\hat{P}x_{n-1} \le x_n$ for all $P \in \mathcal{R}$ and $n \ge 1$ and so $\hat{P}_{k+1} \cdots \hat{P}_n x_0 \le x_{n-k}$ for any $n \ge 1$ and $k < n$. Using this inequality and Lemma 2.6., we find

$$P_1 \cdots P_n x_0(i) \le x_0(i) + \max_{j \in K} x_0(j) + \sum_{k=1}^{n-1} \sum_{h \in K} x_{n-k}(h) \text{ for all } n \ge 1 \text{ and } i \in I.$$

Together this inequality, the finiteness of K and the relations (3.18)-(3.19) imply the Lemma.

We now state our final result.

THEOREM 3.4. *Suppose that the assumptions* 1-4 *hold. Let* $f^{(\infty)}$ *be any stationary policy such that the action* f(i) *minimizes the right side of* (3.12) *for all* $i \in I$. *Then*

$$g = \lim_{n\to\infty} \phi_n(i, f^{(\infty)}) \le \liminf_{n\to\infty} \phi_n(i, \pi) \textit{ for any } i \in I \textit{ and policy } \pi.$$

PROOF. For any memoryless policy π represented by the sequence $(P_1, P_2, \ldots)$ in $\mathcal{R}$, we have that $E_\pi\{v(X_n) | X_0 = i\} = P_1 \ldots P_n v(i)$. Since, for some constant c, $|v(i)| \le c(u^*(i) + y^*(i))$ for all $i \in I$, it now follows from Lemma 3.3. that for any memoryless policy π

(3.20) $$\lim_{n\to\infty} \frac{1}{n} E_{\pi}\{|v(X_n)||X_0=i\} = 0 \text{ for all } i\epsilon I$$

and so, by (1.3) we have that (3.20) holds for any policy π. Now, by observing that we may replace τ_k by $\tau(X_k,a_k)$ in (1.2) a repetition of the well-known proof of Theorem 7.6. in [14] gives the desired result.

REFERENCES

1. Blackwell, D., (1967), Positive dynamic programming, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* Vol. 1, 415-418.
2. Derman, C. and R. Strauch, (1966), A note on memoryless rules for controlling sequential control processes, *Ann. Math. Statist.* 37, 276-278.
3. Derman, C. (1966), Denumerable state Markovian decision processes-average cost criterion, *Ann. Math. Statist.* 37, 1545-1554.
4. Federgruen, A. and H.C. Tijms (1978), The optimality equation in average cost denumerable state semi-Markov decision problems, recurrency conditions and algorithms, *J. Appl. Prob.* 15, 356-373.
5. Federgruen, A., Hordijk, A. and H.C. Tijms (1978), Recurrence conditions in denumerable state Markov decision processes, in: *Dynamic Programming and its Applications,* ed. M.L. Puterman, Academic Press, New York.
6. Federgruen, A., Hordijk, A. and H.C. Tijms (1978), A note on simultaneous recurrence conditions on a set of denumerable stochastic matrices, *J. Appl. Prob.* 15 (to appear).
7. Fisher, L. and S.M. Ross (1968), An example in denumerable decision processes, *Ann. Math. Statist.* 39, 674-675.
8. Hee, K. van, Hordijk, A. and J. van der Wal (1977), Successive approximations for convergent dynamic programming, in: *Markov Decision Theory,* eds. H.C. Tijms and J. Wessels, Mathematical Centre Tract no. 93, Mathematisch Centrum, Amsterdam.
9. Hordijk, A. (1974), *Dynamic Programming and Markov Potential Theory,* Mathematical Centre Tract no. 51, Mathematisch Centrum, Amsterdam.
10. Hordijk, A. (1976), Regenerative Markov decision models, in: Mathematical Programming Study, 6, 49-72, North-Holland, Amsterdam.
11. Maitra, A. (1968), Discounted dynamic programming on compact metric spaces, *Sankhya Ser. A.* 30, 211-216.
12. Robinson, D.R. (1976), Markov decision chains with unbounded costs and applications to the control of queues, *Adv. Appl. Prob.* 8, 159-176.
13. Ross, S.M. (1968), Non-discounted denumerable Markov decision models, *Ann. Math. Statist.* 39, 412-423.
14. Ross, S.M. (1970), *Applied Probability Models with Optimization Applications,* Holden-Day Inc., San Francisco.

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

16. Schäl, M. (1970), Ein verallgemeinertes stationäres entscheidungs modell der dynamische optimierung, in: *Methods of Operations Research,* 10, 145-162, Verlag Anton Hain, Meisenheim.

17. Schäl, M. (1977), On negative dynamic programming with irreducible Markov chains and the average cost criterion, in: *Dynamische Optimierung,* Vol. 98, 93-97, Bonner Mathematische Schriften, Bonn.

18. Taylor, H.M. (1965), Markovian sequential replacement processes, *Ann. Math. Statist.* 36, 1677-1694.

19. Wijngaard, J. (1977), Stationary Markovian decision problems and perturbation theory of quasi-compact linear operators, *Math. of Oper. Res.* 2, 91-102.