stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 17/73 JANUARY

A. HORDIJK and H.C. TIJMS

THE ASYMPTOTIC BEHAVIOUR OF THE MINIMAL TOTAL EXPECTED COST IN DENUMERABLE STATE DYNAMIC PROGRAMMING AND AN APPLICATION IN INVENTORY THEORY

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), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

0. Summary.

In this paper we give for the denumerable state dynamic programming model a set of rather weak conditions under which the minimal total expected cost in the N-stage dynamic programming model minus N times the minimal long-run average expected cost per unit time has a finite limit for each initial state. As an application we prove a conjecture of D.L. Iglehart for the classical dynamic inventory model.

1. *Introduction.*

In this paper we are concerned with the asymptotic behaviour of the minimal total expected cost in denumerable state dynamic programming and with an application in inventory theory. It is shown that, under certain conditions, the minimal total expected cost in the N-stage dynamic programming model minus N times the minimal long-run average expected cost per unit time has a finite limit for each initial state. This was proved for the finite state dynamic programming model by E. Lanery [11] and P.J. Schweitzer [14, 16]. Our proof is an adaptation of these proofs. In [1] and [2] related work has been done for the finite state dynamic programming model. *)

The above result in denumerable state dynamic programming is used to prove a conjecture of D.L. Iglehart [8]. For the classical dynamic inventory model it will be demonstrated that, under the condition of a positive demand, the minimal total expected cost in the N-period inventory model minus N times the minimal average expected cost per period in the infinite period inventory model has a finite limit which can be explicitly given up to a constant. This result was first proved by Iglehart [8] for the case of no set-up cost and was offered as a conjecture for the case of a positive set-up cost.

In section 2 the denumerable state dynamic programming model will be treated and in section 3 the application in inventory theory will be given.

2. *The asymptotic behaviour of the minimal total expected cost in denumerable state dynamic programming.*

We are concerned with a dynamic system which at times t = 1,2,... is observed to be in one of a possible number of states. Let I denote the set of all possible states. We assume I to be *denumerable*. If at time t the

*) In [2] the induction argument used in the proof of lemma 4.7 is incorrect; it seems that this proof cannot be repaired.

system is observed in state i then a decision a must be chosen from a given *finite* set $A(i)$. If the system is in state i at time t and decision a is chosen, then, regardless of the history of the system, two things happen:

(i) we incur an (expected) cost $c(i,a)$ and

(ii) at time t+1 the system will be in state j with probability $p_{ij}(a)$.

The costs $c(i,a)$ and the transition probabilities $p_{ij}(a)$ are assumed to be known. We suppose that the costs $c(i,a)$ are *non-negative*. No further boundedness condition is imposed on the costs.

A *policy* R for controlling the system is any prescription for taking decisions at each point of time. We shall permit a policy for taking a decision at time t to be a function of the entire "history" of the system up to time t. Denote by C the class of all possible policies. A stationary policy, to be denoted by f, is a function which adds to each state $i \in I$ a single decision $f(i) \in A(i)$, such that f prescribes decision $f(i)$ whenever the system is in state i. Given an initial state i and a policy R, denote by X_t and Δ_t, $t = 1,2,\ldots$ the sequences of states and decisions. If a stationary policy f is used, then the sequence of states $\{X_n, n \geq 1\}$ is a Markov chain with transition probabilities $p_{ij}(f) = p_{ij}(f(i))$. Denote by $p_{ij}^{(n)}(f)$ the n-step transition probabilities of this Markov chain, and let

$$\pi_{ij}^{(n)}(f) = \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)}(f) \qquad \text{for } n = 1,2,\ldots \text{ and } i,j \in I \; . \tag{2.1}$$

It is well-known from Markov chain theory that [3]

$$\pi_{ij}(f) = \lim_{n\to\infty} \pi_{ij}^{(n)}(f) \qquad \text{exists for all } i,j \in I \; , \tag{2.2}$$

where $\sum_{j\in I} \pi_{ij}(f) \leq 1$ for all $i \in I$.

For any $i \in I$ and any policy $R \in C$ let,

$$\phi(i,R) = \liminf_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E_R\{c(X_t,\Delta_t)|X_1 = i\} \; , \tag{2.3}$$

where E_R denotes the expectation under policy R. Observe that $\phi(i,R)$ exists ($+\infty$ is admitted), since the costs $c(i,a)$ are non-negative. When the limit exists $\phi(i,R)$ represents the long-run average expected cost per unit time when the initial state is i and policy R is used. A policy $R^* \in C$ is said to be *average cost optimal* if $\phi(i,R^*) = \inf_{R\in C} \phi(i,R)$ for all $i \in I$.

We shall now introduce a number of assumptions.

Assumption 1. There is a set of finite numbers $\{v(i),\ g \mid i \in I\}$ *such that* $\sum_{j\in I} p_{ij}(a)v(j)$ *is absolutely convergent for all* $a \in A(i)$ *and all* $i \in I$,

$$v(i) = \min_{a\in A(i)} \{c(i,a) - g + \sum_{j\in I} p_{ij}(a)v(j)\} \quad \textit{for all } i \in I \tag{2.4}$$

and

$$\lim_{n\to\infty} \frac{1}{n} E_R\{v(X_n)|X_1 = i\} = 0 \qquad \textit{for all } i \in I \textit{ and all } R \in C . \tag{2.5}$$

Define the class F_{opt} of stationary policies as

$$F_{opt} = \{f \mid f \text{ is a stationary policy such that } f(i) \text{ minimizes the right-hand side of (2.4) for each } i \in I\} . \tag{2.6}$$

By the remark following the proof of theorem 1 in [12] ,

$$g = \inf_{R\in C} \phi(i,R) \qquad \text{for all } i \in I \tag{2.7}$$

and

$$\phi(i,f) = g \qquad \text{for all } i \in I \text{ and all } f \in F_{opt}. \tag{2.8}$$

Hence the minimal long-run average expected cost per unit time is independent of the initial state. Moreover, each stationary policy $f \in F_{opt}$ is average cost optimal.

Remark 2.1

It can be shown that under rather general conditions $\inf_{R \in C} \phi(i,R)$ is independent of the initial state i. For the case of a finite state space I it is proved in [1] that if there is stationary policy such that I is a positive class under that policy, then the minimal long-run average expected cost per unit time is independent of the initial state. This result has been proved in [5] for a denumerable state space I under the condition that for each pair of states i and j there is a stationary policy such that i and j are positive recurrent under that policy and belong to a same positive class. However, if $\inf_{R \in C} \phi(i,R)$ is bounded, then this condition can be considerably weakened. In this case it is sufficient to require that for each pair of states i and j holds $\sup_{R \in C} \alpha_R(i,j) = 1$, where $\alpha_R(i,j)$ is the probability that state j will be ever reached when the initial state is i and policy R is used.

Together with the well-known fact that Howard's policy improvement method [6] leads to a solution of the optimality equation when I is finite, the above results imply that assumption 1 is certainly satisfied in the case where I is finite and for each pair of states i and j there is a policy $R \in C$ such that state j will be reached with probability one when the initial state is i and policy R is used.

Assumption 2. For each stationary policy f the associated Markov chain $\{X_n\}$ is non-dissipative, that is, from each initial state the set of positive recurrent states will be reached with probability one.

The Markov chain $\{X_n\}$ associated with a stationary policy f is non-dissipative if and only if $\sum_{j \in I} \pi_{ij}(f) = 1$ for all $i \in I$ [3].

Assumption 3. For each policy $f \in F_{opt}$ holds that each state which is positive recurrent under policy f is aperiodic.

Assumption 4. For each average cost optimal stationary policy the associated Markov chain $\{X_n\}$ has no two disjoint closed sets.

To introduce the last assumption, we fix an arbitrary finite function $v_0(i)$ such that

$$(2.9) \qquad \sum_{j\in I} p_{ij}(a)v_0(j) \text{ is finite and is bounded from below for } a\in A(i) \text{ and } i\in I,$$

and we define the sequence of functions $v_n(i)$, $i \in I$, by

$$(2.10) \qquad v_n(i) = \min_{a\in A(i)} \{c(i,a) + \sum_{j\in I} p_{ij}(a)v_{n-1}(j)\} \text{ for } i \in I;\ n = 1,2,\ldots$$

Observe that, by (2.9) and $c(i,a) \geq 0$, the function $v_n(i)$ exists. If we suppose that in the N-stage dynamic programming model a salvage cost $v_0(j)$ is incurred when the final state is j, then $v_N(i)$ can be interpreted as the minimal total expected cost in the N-stage dynamic programming model when the initial state is i (cf. [4]).

Assumption 5. The function $v_1(i) - v(i)$, $i \in I$, is bounded.

It will appear (p.12) that together the assumptions 1 and 5 imply $\sum_j p_{ij}(a)v_{n-1}(j)$ is absolutely convergent for all i,a and n, and so, the function $v_n(i)$, $i \in I$, is finite for all n.

Under the assumptions 1,2,3 and 5,

$$(2.11) \qquad \lim_{n\to\infty} \{v_n(i) - ng - v(i)\} \quad \text{is finite for each } i \in I \text{ and is bounded.}$$

If in addition assumption 4 holds, then, for some finite constant c,

$$(2.12) \qquad \lim_{n\to\infty} \{v_n(i) - ng\} = v(i) + c \qquad \text{for all } i \in I\ .$$

These limit relations can be established by adapting proofs given by

E. Lanery [11] and P.J. Schweitzer [16] for the case of a finite state space. This will be done in the appendix.

3. A limit theorem in inventory theory.

We consider an inventory model in which the demands $\xi_1, \xi_2, \ldots$ for a single item in periods $t = 1,2,\ldots$ are independent and non-negative random variables having a common discrete probability distribution $\phi(j) = P\{\xi_t = j\}$, $(j = 0,1,\ldots;\ t = 1,2,\ldots)$. It is assumed that $\mu = E\xi_t$ is finite. At the beginning of each period the stock on hand is reviewed. At each review an order may be placed for any positive integral amount of stock. An order, when placed, is immediately delivered. The demand in each period takes place after review and delivery (if any). Furthermore, we assume that any unfilled demand in a period is completely backlogged to be eventually satisfied by future deliveries. Hence the stock on hand may take on negative values indicating the existence of a backlog. The stock on hand may take on any integral value. The following costs are involved. The cost of ordering j units is $K\delta(j) + cj$, where $K \geq 0$, $c \geq 0$, $\delta(0) = 0$ and $\delta(j) = 1$ for $j \geq 1$. Let $L(k)$ be the holding and shortage cost in a period when k is the amount of stock on hand at the beginning of that period just after any additions to stock. It is assumed that $L(k)$ is *convex*, i.e. $L(k+1) - L(k) \geq L(k) - L(k-1)$ for each integer k. Moreover, we assume $L(k) \geq 0$ and $L(k) \to \infty$ as $|k| \to \infty$.

We shall demonstrate for this model that, under the condition of a positive demand, the minimal total expected cost in the N-period model minus N times the minimal average expected cost in the infinite period model has a finite limit for each initial state. This limit fuction can be explicitly given up to a constant. This result was first proved by D.L. Iglehart [8] for the case of $K = 0$ and was offered as a conjecture for the case of $K > 0$. In his paper Iglehart assumes a continuous positive demand.

To prove the above result, we first give a number of known optimality results for this inventory model.

A) The finite period model. Let $\mathbb{Z}$ be the set of all integers. Define

$$(3.1) \qquad v_0(i) = 0 \qquad \text{for each } i \in \mathbb{Z}$$

and for n = 1,2,... , let

$$(3.2) \qquad v_n(i) = \inf_{k \geq i} \{c.(k-i) + K\delta(k-i) + L(k) + \sum_{j=0}^{\infty} v_{n-1}(k-j)\phi(j)\}, \; i \in \mathbb{Z}$$

The choice $v_0(i) \equiv 0$ can be interpreted as follows. In the finite period model it is assumed that stock left over at the end of the final period has no value and backlogged demand remaining at the end of the final period is satisfied at a cost zero. Another choice of $v_0(i)$ will be considered in the remark at the end of this section. The quantity $v_N(i)$ is the minimal total expected cost in the N-period model when the initial stock is i. Moreover, a famous proof due to Scarf [13] shows that, for each n = 1,2,... ,

$$(3.3) \qquad v_n(i) = \begin{cases} -ci + K + G_n(S_n) & \text{for } i < s_n , \\ -ci + G_n(i) & \text{for } i \geq s_n , \end{cases}$$

where S_n is an integer which minimizes the finite (K-convex) function

$$(3.4) \qquad G_n(k) = ck + L(k) + \sum_{j=0}^{\infty} v_{n-1}(k-j)\,\phi(j) , \qquad k \in \mathbb{Z}$$

and $s_n (\leq S_n)$ is the unique integer satisfying $G_n(s_n) \leq K + G_n(S_n) < G_n(s_n - 1)$. Hence we can replace inf by min in (3.2). The right-hand side of (3.2) is minimal for $k = S_n$ when $i < s_n$ and for $k = i$ when $i \geq s_n$. For the N-period model the following policy of the (s,S) type achieves the minimal total expected cost $v_N(i)$: If at the beginning of period t the stock on hand $i < s_t$, order $S_t - i$ units; otherwise, do not order in period t, (t=1,...,N).

Finally, we mention that the integers s_n and S_n are bounded for $n \geq 1$

[7,8,10,18]. This important result will also be needed in our analysis.

B) The infinite period model. Let us first introduce some notation. Denote by $\phi^{(n)}(j)$ the n-fold convolution of the probability distribution $\phi(j)$ with itself, and let

$$(3.5) \qquad m(j) = \sum_{n=1}^{\infty} \phi^{(n)}(j) \text{ and } M(j) = \sum_{k=0}^{j} m(k) \text{ for } j = 0,1,\dots .$$

The renewal function $M(j)$ is finite and the numbers $m(j)$ can be computed from

$$(3.6) \qquad m(j) = \phi(j) + \sum_{k=0}^{j} \phi(j-k)m(k) \qquad \text{for } j = 0,1,\dots .$$

Let s and S be any two integers with $s \leq S$. A stationary policy which is frequently used in inventory problems is the familiar (s,S) policy, that is, if, at review, the stock on hand $i < s$, then $S-i$ units are ordered; otherwise, no order is placed. Under an (s,S) policy the long-run average (expected) cost per period is given by [8,17,18]

$$a(s,S) = \frac{L(S) + \sum_{k=0}^{S-s} L(S-k)m(k) + K}{1 + M(S-s)} + c\mu ,$$

independent of the initial stock. Let g be defined as

$$(3.7) \qquad g = \min\{a(s,S) \mid s \leq S,\ s,S \in \mathbf{Z}\}$$

The constant g exists and is finite. *Fix* now two finite integers s^* and S^* with $s^* \leq S^*$ such that

$$(3.8) \qquad g = a(s^*,S^*) \text{ and } \qquad L(s^*-1) \geq g-c\mu \geq L(s^*) .$$

It is known that such integers exist [8,9,17]. From definition the (s^*,S^*) policy is optimal with respect to the average cost criterion among the class of the (s,S) policies. However, the (s^*,S^*) policy is also average cost optimal among the class of all possible policies [8,9,17]. Moreover, we note that if $\phi(1) > 0$, then the first part of (3.8) implies the second part of (3.8) [9,17].

Define now the finite function $v(i)$, $i \in \mathbb{Z}$, as

$$(3.9) \qquad v(i) = \begin{cases} -c.(i-s^*+1) & \text{for } i < s^* . \\ L(i) + \sum_{k=0}^{i-s^*} L(i-k)m(k) - (g-c\mu)\,\{1+M(i-s^*)\} & \text{for } i \geq s^* . \end{cases}$$

Then [8,17]

$$(3.10) \qquad v(i) = \min_{k \geq i}\{c.(k-i)+K\delta(k-i)+L(k)+\sum_{j=0}^{\infty} v(k-j)\phi(j)\} \qquad \text{for } i \in \mathbb{Z} ,$$

where the right-hand side of (3.10) is minimized by $k = S^*$ for $i < s^*$ and by $k = i$ for $i \geq s^*$.

We are now ready to prove that, for some finite constant γ ,

$$(3.11) \qquad \lim_{n\to\infty} \{v_n(i) - ng\} = v(i) + r \qquad \text{for all } i \in \mathbb{Z} ,$$

provided that the following assumption is satisfied:

Assumption. There is a finite integer r such that $\phi(i) > 0$ for all $i \geq r$

To prove this, we shall define a Markovian decision model which has the same probabilistic structure and the same cost structure as the inventory model under consideration. Fix two finite integers L and U such that

$$(3.12)\qquad L < s^*,\ L < s_n \text{ and } U > S^* \text{ and } U > S_n \qquad \text{for all } n = 1,2,\ldots$$

These integers L and U can be chosen, since s_n and S_n are bounded. Consider now a Markovian decision model from which the state space I, the set $A(i)$ of possible decisions in state i, the costs $c(i,a)$ and the transition probabilities $p_{ij}(a)$ are given by

$$(3.13)\quad I=\{i \mid i \text{ integer},\ i\le U\},\ A(i)=\{a \mid a \text{ integer},\ a\ge i \text{ and } L\le a\le U\},\ i \in I,$$

$$(3.14)\quad c(i,a) = c.(a-i)+K\delta(a-i)+L(a) \text{ and } p_{ij}(a) = \phi(a-j),\quad a\in A(i);\ i,j\in I\ ,$$

where we define $\phi(k) = 0$ for $k < 0$. Since the right-hand side of (3.2) is minimized by $k = S_n$ for $i < s_n$ and by $k = i$ for $i \geq s_n$ and since the right-hand side of (3.10) is minimized by $k = S^*$ for $i < s^*$ and by $k = i$ for $i \geq s^*$, it is easily verified from (3.2), (3.10),(3.12), (3.13) and (3.14) that

$$(3.15)\qquad v_n(i) = \min_{a\in A(i)} \{c(i,a)+ \sum_{j\in I} p_{ij}(a)v_{n-1}(j)\} \text{ for } i\in I \text{ and } n=1,2,\ldots\ ,$$

where $v_0(i) = 0$ for all $i \in I$, and

$$(3.16)\qquad v(i) = \min_{a\in A(i)} \{c(i,a) - g + \sum_{j\in I} p_{ij}(a)v(j)\} \quad \text{for all } i \in I\ .$$

We shall now verify the assumptions 1-5 in section 2. Let us first check assumption 1. We have already shown that the finite numbers g and v(i) defined by (3.7) and (3.9) satisfy (2.4) (see(3.16)). Since v(i) is linear for $i < s^*$ and $\mu = \sum j\phi(j)$ is finite, it follows that $\sum_{j=0}^{\infty} v(a-j)\phi(j)$ is absolutely convergent for all $a\in A(i)$. To prove (2.5), we note that X_t represents the stock on hand just before ordering in period t and Δ_t represents the stock on hand just after ordering in period t. Since excess demand is backlogged, we have

$$(3.17) \qquad X_{t+1} = \Delta_t - \xi_t \qquad \text{for all } t = 1,2,\ldots$$

By the choice of I and $A(i)$,

$$(3.18) \qquad X_t \le U \text{ and } L \le \Delta_t \le U \qquad \text{for all } t = 1,2,\ldots \, .$$

Since $v(i)$ is linear for $i < s^*$ and $E\xi_t = \mu < \infty$, it now follows easily from (3.17) and (3.18) that $E_R\{v(X_n) \mid X_1 = i\}$ is bounded in n and R, and so, (2.5) holds. This proves assumption 1. By (3.3) and (3.9) the function $v_1(i) - v(i)$ is bounded in $i \in I$ (observe we need $i \le U$), and so, assumption 5 holds. So far we have not used the assumption that $\phi(i)>0$ for all $i \ge r$ for some r. However, this assumption and the choice of I and $A(i)$ for $i \in I$ imply that for each stationary policy the associated Markov chain $\{X_n\}$ has a class of aperiodic positive recurrent states, has only a finite number of transient states and has no two disjoint closed sets. Hence the assumptions 2,3 and 4 are also satisfied, and so, the assumptions 1-5 hold. Since $I = \{i \mid i \le U\}$, it now follows from (2.12) that, for some finite constant γ,

$$(3.19) \qquad \lim_{n\to\infty} \{v_n(i) - ng\} = v(i) + \gamma \qquad \text{for all } i \le U \, .$$

This constant γ is independent of U, since U can be chosen arbitrarily large. Letting $U \to\infty$ in (3.19), we obtain the result (3.11).

Remark 3.1. In the foregoing discussion we have assumed that $v_0(i) = 0$ for all i. Another interesting choice is $v_0(i) = -ci$ for all $i \in \mathbb{Z}$. This choice can be interpreted as follows. In the finite period inventory model it is assumed that each unit of stock left over at the end of the final period can be salvaged with a return of c and each unit of backlogged demand remaining at the end of the final period is satisfied at a cost c. For this model, denote by $v'_N(i)$ the minimal total expected cost in the N-period model when the initial stock is i. Using the fact that this model with a salvage value c and a salvage cost c can be reduced to an equivalent model with a salvage value zero and a salvage cost zero (cf. [18, pp. 528-529]), it is easily shown that, for some finite constant β,

$$\lim_{n\to\infty} \{v'_n(i) - ng\} = v(i) + \beta \qquad \text{for all } i \in Z ,$$

where, of course, g and v(i) are given by (3.7) and (3.9).

Appendix.

In this appendix we shall prove the limit theorem given in section 2. Our proof is an adaptation of proofs given by E. Lanery [11] and P.J. Schweitzer [16] for the case of a finite state space I.

Lemma 1. Suppose the assumptions 1 and 5 are satisfied. Then there is a finite constant N such that $|v_n(i)-ng-v(i)| \le N$ for all $n = 1,2,\ldots$ and $i \in I$.

Proof. By assumption 5, there is a finite constant, say N, such that $v_1(i)-g-v(i)$ is bounded by N. Assume that we have shown that $|v_k(i)-kg-v(i)| \le N$ for $1 \le k \le n$ and all $i \in I$. From this and assumption 1, $\sum_j p_{ij}(a)|v_n(j)| < \infty$ for all i and a. Let $f \in F_{opt}$, then

(1) $$v(i) = c(i,f(i)) - g + \sum_{j\in I} p_{ij}(f)v(j) \qquad \text{for all } i \in I .$$

From (2.10),

(2) $$v_{n+1}(i) \le c(i,f(i)) + \sum_{j\in I} p_{ij}(f)v_n(j) \qquad \text{for all } i \in I .$$

By (1) and (2),

(3) $$v_{n+1}(i)-(n+1)g-v(i) \le \sum_{j\in I} p_{ij}(f)\{v_n(j)-ng-v(j)\} \quad \text{for all } i \in I .$$

By the induction hypothesis we have $v_n(j)-ng-v(j) \le N$, and so

(4) $$v_{n+1}(i)-(n+1)g-v(i) \le N \qquad \text{for all } i \in I .$$

To prove that $v_{n+1}(i)-(n+1)g-v(i) \ge -N$, let a_i be such that (see(2.10))

(5) $$v_{n+1}(i) = c(i,a_i) + \sum_{j\in I} p_{ij}(a_i)v_n(j) \qquad \text{for all } i \in I .$$

Then, by (2.4),

(6) $$v(i) \le c(i,a_i)-g+ \sum_{j\in I} p_{ij}(a_i)v(j) \qquad \text{for all } i \in I .$$

From (5), (6) and the induction hypothesis

(7) $$v_{n+1}(i)-(n+1)g-v(i) \ge \sum_{j\in I} p_{ij}(a_i)\{v_n(j)-ng-v(j)\}\ge -N \text{ for all } i \in I$$

which proves the lemma.

Theorem 1. Suppose the assumptions 1 and 5 are satisfied. Let $f \in F_{opt}$ and assume that K is a positive recurrent class with aperiodic states for the Markov chain $\{X_n\}$ associated with the stationary policy f. Then

(8) $$\lim_{n\to\infty} \{v_n(i)-ng-v(i)\} \text{ exists and is finite for all } i \in K \text{ and, moreover, is independent of } i \in K.$$

Proof. For $n = 0,1,\ldots$ and $i \in I$, let

(9) $$v_n^*(i) = v_n(i) - ng - v(i) .$$

The sequence $\{v_n^*(i),\ n \geq 1\}$ is bounded for each $i \in I$, by lemma 1. We have to prove that for each $i \in K$ the sequence $\{v_n^*(i)\}$ has only one limit point and that for each $i \in K$ this limit point has the same value. To do this, we fix an arbitrary state $r \in K$. Let α and β be two limit points of $\{v_n^*(r),\ n \geq 1\}$. By the well-known diagonalization method and the boundedness of the sequences $\{v_n^*(j),\ n \geq 1\}$, we can get two sequences $\{n_k, k \geq 1\}$ and $\{m_h, h \geq 1\}$ with $n_k \to \infty$ and $m_h \to \infty$ such that for all $i \in I$,

(10) $$\lim_{k\to\infty} v_{n_k}^*(i) \text{ exists and is equal to } x_i \text{ (say), where } x_r = \alpha$$

and

(11) $$\lim_{h\to\infty} v_{m_h}^*(i) \text{ exists and is equal to } y_i \text{ (say), where } y_r = \beta\ .$$

By lemma 1, the numbers x_i and y_i, $i \in I$ are bounded. In lemma 1 we have proved that (see(3)),

(12) $$v_{n+1}^*(i) \leq \sum_{j\in I} p_{ij}(f)v_n^*(j) \qquad \text{for all } n=1,2,\ldots \text{ and all } i \in I\ .$$

By applying (12) repeatedly and using the boundedness of $\{v_n^*(j), j \in I\}$, we get

(13) $$v_{n+m}^*(i) \leq \sum_{j\in I} p_{ij}^{(m)}(f)v_n^*(j) \qquad \text{for all } n,m=1,2,\ldots \text{ and all } i \in I.$$

Since K is a positive recurrent class under policy f, we have $p_{ij}^{(n)}(f) = 0$ for all $i \in K$ and $n \geq 1$ when $j \notin K$. Hence

(14) $$v_{n+m}^*(i) \leq \sum_{j\in K} p_{ij}^{(m)}(f)v_n^*(j) \qquad \text{for all } n,m=1,2,\ldots \text{ and all } i \in K.$$

Since the states of the positive recurrent class K are aperiodic, it follows from Markov chain theory [3] that, for each $i,j \in K$, $p_{ij}^{(n)}(f)$ has a limit which is independent of $i \in K$. For any $i,j \in K$, let

$$(15) \qquad \pi_j(f) = \lim_{n\to\infty} p_{ij}^{(n)}(f), \text{ then } \pi_j(f)>0 \text{ for all } j\in K \text{ and } \sum_{j\in K} \pi_j(f)=1 .$$

We shall now prove that

$$(16) \qquad y_i \leq \sum_{j\in K} \pi_j(f)x_j \qquad \text{for all } i \in K .$$

To do this, we choose for each integer $k \geq 1$ a positive integer $h(k)$ such that $m_{h(k)}-n_k > k$. Let $s_k = m_{h(k)} - n_k$ for $k \geq 1$, then $s_k \to\infty$ as $k \to\infty$. By (14),

$$(17) \qquad v^*_{n_k+s_k}(i) \leq \sum_{j\in K} p_{ij}^{(s_k)}(f)v^*_{n_k}(j) \quad \text{for all } k \geq 1 \text{ and all } i \in K .$$

Since $\{n_k+s_k, k \geq 1\}$ is a subsequence of $\{m_h, h \geq 1\}$, we have by (11) that

$$(18) \qquad \lim_{k\to\infty} v^*_{n_k+s_k}(i) = y_i \qquad \text{for all } i \in I .$$

Since K is a positive recurrent class, we have by theorem 4 on p. 37 in [3] that for each $i \in K$ the series $\sum_{j\in K} p_{ij}^{(n)}(f)$ converges uniformly with respect to n. Thus for each $\varepsilon > 0$ and each $i \in K$ there is a finite subset $J = J(i,\varepsilon)$ such that

$$(19) \qquad \sum_{j\in K\setminus J} p_{ij}^{(n)}(f) \leq \varepsilon \qquad \text{for all } n = 1,2,\ldots$$

It now follows easily from the boundedness of $v_n^*(j)$ (see lemma 1), (10), (15) and (19) that

$$(20) \qquad \lim_{k\to\infty} \sum_{j\in K} p_{ij}^{(s_k)}(f)v^*_{n_k}(j) = \sum_{j\in K} \pi_j(f)x_j \qquad \text{for all } i \in K .$$

From (17) and (20) follows (16). In the same way we can prove that

(21) $$x_i \leq \sum_{j\in K} \pi_j(f)y_j \qquad \text{for all } i \in K .$$

Substituting (21) in (16) and (16) in (21) and using $\sum_j \pi_j(f) = 1$, we get

(22) $$y_i \leq \sum_{j\in K} \pi_j(f)y_j \text{ and } x_i \leq \sum_{j\in K} \pi_j(f)x_j \qquad \text{for all } i \in K .$$

Multiplying both sides of each inequality in (22) by $\pi_i(f)$, taking the sum on $i \in K$ and using (15), we see that for each $i \in K$ the equality signs must hold in (22). From this, (16) and (21) we get

(23) $$x_i \leq \sum_{j\in K} \pi_j(f)y_j = y_i \leq \sum_{j\in K} \pi_j(f)x_j = x_i \qquad \text{for all } i \in K ,$$

which shows that, for some finite constant c,

(24) $$x_i = y_i = c \qquad \text{for all } i \in K .$$

In particular $x_r = y_r$, and so, $\alpha = \beta$. Hence the sequence $\{v_n^*(r), n \geq 1\}$ has only one limit point, and so, this sequence is convergent. However, since r was arbitrarily chosen in K, it follows that for each $i \in K$ the sequence $\{v_n^*(i)\}$ is convergent. Since x_i is a limit point of $\{v_n^*(i)\}$, it follows that for each $i \in K$ the sequence $\{v_n^*(i)\}$ has the finite limit x_i and, by (24), this limit is the same for all $i \in K$. This ends the proof.

Remark 1. Suppose the assumptions 1 and 5 are satisfied. Let $f \in F_{opt}$ and let K be a positive recurrent class of period d under policy f. For any $i \in K$, let the subclass $K_d(i)$ be defined as $K_d(i) = \{j \mid p_{ij}^{(nd)}(f) > 0$ for some $n \geq 1\}$. Then (cf. [3]), for any $i \in K$,

(25) $$\lim_{n\to\infty} p_{ij}^{(nd)}(f) = d\pi_j(f) \text{ for } j\in K_d(i), \text{ and } \sum_{j\in K_d(i)} \pi_j(f) = \frac{1}{d} ,$$

where $\pi_j(f) = \lim_{n\to\infty} \pi_{ij}^{(n)}(f)$ independent of $i \in K$ for all $j \in K$ and $\pi_{ij}^{(n)}(f)$ is defined by (2.1). It now follows easily from an examination of the proof of theorem 1 that in the periodic case

(26) $$\lim_{n\to\infty} \{v_{nd+s}(i)-(nd+s)g-v(i)\} \text{ is finite for all } i\in K \text{ and all } s=1,\ldots,d.$$

Moreover, this limit is independent of i within a given subclass.

Lemma 2. Suppose the assumption 1 is satisfied. Let f be a stationary policy such that for each state i which is positive recurrent under f,

(27) $$v(i) = c(i,f(i)) - g + \sum_{j\in I} p_{ij}(f)v(j) .$$

Then there is a policy $f^* \in F_{opt}$ such that each state which is positive recurrent under policy f is also positive recurrent under policy f^*.

Proof. The proof is quite simple. Let $f^*(i) = f(i)$ for each state i which is positive recurrent under f, and, for the other states, let $f^*(i)$ be a decision which minimizes the right-hand side of (2.4) in assumption 1. By this construction, $f^* \in F_{opt}$ and f and f^* are identical on each positive recurrent class of f, and so, each state which is positive recurrent under f is also positive recurrent under f^*.

The next lemma has been proved in [15] for the case of a finite I.

Lemma 3. Suppose the assumptions 1, 2 and 4 are satisfied. Let $\beta(i)$, $i \in I$, be a bounded function such that

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

Then, for some finite constant β, $\beta(i) = \beta$ for all $i \in I$.

Proof. Let us first note that, by (2.5), $\lim_{n\to\infty}(1/n)\ E_R\{v(X_n)+\beta(X_n)\,|\,X_1=i\}=0$ for all $i \in I$ and all $R \in C$. Let h be a stationary policy such that for each $i \in I$ the decision $h(i)$ minimizes the right-hand side of (28). By the remark following theorem 1 in [12], we have that $\phi(i,h) = g$ for all $i \in I$. Hence the stationary policy h is optimal with respect to the average cost criterion. Choose a stationary policy $f \in F_{opt}$. By the definitions of f and h, we have

(29) $$v(i) = c(i,f(i)) - g + \sum_{j\in I} p_{ij}(f)v(j) \qquad \text{for all } i \in I$$

(30) $$v(i) \leq c(i,h(i)) - g + \sum_{j\in I} p_{ij}(h)v(j) \qquad \text{for all } i \in I\ ,$$

(31) $$v(i)+\beta(i) = c(i,h(i))-g+ \sum_{j\in I} p_{ij}(h)\{v(j)+\beta(j)\} \text{ for all } i \in I\ ,$$

(32) $$v(i)+\beta(i) \leq c(i,f(i))-g+ \sum_{j\in I} p_{ij}(f)\{v(j)+\beta(j)\} \text{ for all } i \in I\,.$$

Observe that each series in (29)-(31) is absolutely convergent. By (29) and (32),

(33) $$\beta(i) \leq \sum_{j\in I} p_{ij}(f)\beta(j) \qquad \text{for all } i \in I\,.$$

Iterating (33) and using the boundedness of $\beta(j)$, we get $\beta(i) \leq \sum_{j\in I} p_{ij}^{(n)}(f)\beta(j)$, and so,

(34) $$\beta(i) \leq \sum_{j\in I} \pi_{ij}^{(n)}(f)\beta(j) \qquad \text{for all } n \geq 1 \text{ and all } i \in I\ .$$

By assumption 2, $\sum_{j\in I} \pi_{ij}(f) = 1$ for all $i \in I$. This implies that $\sum_{j\in I} \pi_{ij}^{(n)}(f)$ converges uniformly with respect to n for each $i \in I$ [3,p.37]. Using this and the boundedness of $\beta(j)$, we get by letting $n\to\infty$ in (34) that

(35) $$\beta(i) \leq \sum_{j \in I} \pi_{ij}(f)\beta(j) \qquad \text{for all } i \in I \text{ .}$$

Denote by R(f) respectively R(h) the set of states which are positive recurrent under f respectively h. Since both f and h are average cost optimal, we have by assumption 4 that $\pi_{ij}(f) = \pi_j(f)$ and $\pi_{ij}(h) = \pi_j(h)$ independently of i. Hence

(36) $$\beta(i) \leq \sum_{j \in I} \pi_j(f)\beta(j) \qquad \text{for all } i \in I \text{ .}$$

By multiplying both sides of (36) with $\pi_i(f)$, taking the sum on i and using that $\pi_i(f) > 0$ for $i \in R(f)$ and $\sum \pi_j(f) = 1$, we get

(37) $$\beta(i) = \sum_{j \in I} \pi_j(f)\beta(j) \qquad \text{for all } i \in R(f) \text{ .}$$

By (30) and (31), $\beta(i) \geq \sum_{j \in I} p_{ij}(h)\beta(j)$ for all $i \in I$. From this we deduce in the same way as above

(38) $$\begin{aligned} \beta(i) &\geq \sum_{j \in I} \pi_j(h)\beta(j) && \text{for all } i \in I \text{ ,} \\ \beta(i) &= \sum_{j \in I} \pi_j(h)\beta(j) && \text{for all } i \in R(h) \text{ .} \end{aligned}$$

It is easily seen that assumption 4 implies that $R(f) \cap R(h)$ is not empty, and so, by (37) and the last part of (38), $\sum_j \pi_j(f)\beta(j) = \sum_j \pi_j(h)\beta(j) = \beta$ (say). Next it follows from (36) and the first part of (38) that $\beta(i) = \beta$ for all $i \in I$. This ends the proof *).

---

*) This proof and the proof of theorem 2 below are the only proofs which need assumption 4. It follows from the above proof that in assumption 4 the condition of no two disjoint closed sets need be imposed only on the average cost optimal stationary policies which are also "functional-optimal".

The proof of the next main theorem is a direct generalisation of a proof due to P.J. Schweitzer [16].

Theorem 2.

(a) Suppose the assumptions 1, 2, 3 and 5 are satisfied. Then

(39) $\lim_{n\to\infty} \{v_n(i)-ng-v(i)\}$ exists for each $i \in I$ and is bounded in $i\in I$.

(b) Suppose the assumptions 1-5 are satisfied. Then the limit function given by (39) is a constant one.

Proof.

(a) Let us recall that $v_n^*(i)$ is defined by

(40) $$v_n^*(i) = v_n(i) - ng - v(i) \qquad \text{for } n = 0,1,\dots \text{ and } i \in I.$$

By (2.10) and the fact that $\sum p_{ij}(a)v_n(j)$ and $\sum p_{ij}(a)v(j)$ are absolutely convergent,

(41) $$v_n^*(i) = \min_{a\in A(i)} \{b(i,a)+\sum_{j\in I} p_{ij}(a)v_{n-1}^*(j)\} \text{ for } n \geq 1 \text{ and } i \in I ,$$

where

(42) $$b(i,a) = c(i,a)-g+\sum_{j\in I} p_{ij}(a)v(j)-v(i) \quad \text{for } a\in A(i) \text{ and } i \in I .$$

It follows from assumption 1 that

(43) $$\min_{a\in A(i)} b(i,a) = 0 \qquad \text{for all } i \in I .$$

Define now

(44) $$m(i)=\liminf_{n\to\infty} v_n^*(i), \quad M(i)=\limsup_{n\to\infty} v_n^*(i) \qquad \text{for all } i \in I .$$

By lemma 1, the sets of numbers $\{m(i),\ i \in I\}$ and $\{M(i),\ i \in I\}$ are bounded. We have to prove that $m(i) = M(i)$ for all $i \in I$. To do this, we shall first show

$$(45) \qquad m(i) \geq \min_{a\in A(i)} \{b(i,a) + \sum_{j\in I} p_{ij}(a)m(j)\} \quad \text{for all } i \in I ,$$

and

$$(46) \qquad M(i) \leq \min_{a\in A(i)} \{b(i,a) + \sum_{j\in I} p_{ij}(a)M(j)\} \quad \text{for all } i \in I .$$

To prove (45), fix $i_0 \in I$. Since for each $i \in I$ the sequence $\{v_n^*(i), n\geq 1\}$ is bounded, we can, by the diagonalization method, get a sequence $\{n_k\}$ with $n_k \to\infty$ such that

$$(47) \qquad \lim_{k\to\infty} v^*_{n_k}(i_0) = m(i_0) \text{ and}$$

$$(48) \qquad \lim_{k\to\infty} v^*_{n_k-1}(i) \text{ exists for all } i \in I \text{ and is equal to } \phi(i) \text{ (say)} .$$

Of course, $\phi(i) \geq m(i)$ for all $i \in I$, since $m(i)$ is the smallest limit point of $\{v_n^*(i)\}$. It follows from lemma 1 and the bounded convergence theorem that

$$(49) \qquad \lim_{k\to\infty} \sum_{j\in I} p_{i_0 j}(a)v^*_{n_k-1}(j) = \sum_{j\in I} p_{i_0 j}(a)\phi(j) \geq \sum_{j\in I} p_{i_0 j}(a)m(j)$$

for all $a \in A(i_0)$. Choose $\varepsilon > 0$. Since $A(i_0)$ is finite, it follows from (47) and (49) that there is an integer k_0 such that for all $a \in A(i_0)$ and all $k \geq k_0$,

$$(50) \qquad v^*_{n_k}(i_0) \leq m(i_0)+\varepsilon \text{ and } \sum_{j\in I} p_{i_0 j}(a)v^*_{n_k-1}(j) \geq \sum_{j\in I} p_{i_0 j}(a)m(j)-\varepsilon.$$

From (41) and (50), for $k \geq k_0$,

$$(51)\quad m(i_0)+\varepsilon \geq v^*_{n_k}(i_0) = \min_{a\in A(i_0)} \{b(i_0,a) + \sum_{j\in I} p_{i_0 j}(a)v^*_{n_k-1}(j)\} \geq$$

$$\geq \min_{a\in A(i_0)} \{b(i_0,a) + \sum_{j\in I} p_{i_0 j}(a)m(j)\}-\varepsilon \ ,$$

from which we get (45) since $\varepsilon > 0$ and i_0 were arbitrarily chosen. In a very similar way we can prove (46).

Let the stationary policy f be defined such that f(i) minimizes the right-side of (45) for each $i \in I$. Then, by (45) and (46),

$$(52)\quad b(i,f(i))+\sum_{j\in I} p_{ij}(f)m(j) \leq m(i)\leq M(i)\leq b(i,f(i)+\sum_{j\in I} p_{ij}(f)M(j),\ i \in I\ .$$

Using $\pi_{kj}(f) = \sum_{i\in I} \pi_{ki}(f)p_{ij}(f)$ for all $i,j \in I$ and using the boundedness of $\{m(i)\}$, it follows from (52) that

$$(53)\qquad \sum_{i\in I} \pi_{ki}(f)b(i,f(i))+\sum_{j\in I} \pi_{kj}(f)m(j)\leq \sum_{i\in I} \pi_{ki}(f)m(i) \text{ for all } k \in I.$$

Observe that the first series in (53) is defined because $b(i,a)\geq 0$. From (53),

$$(54)\qquad \sum_{j\in I} \pi_{ij}(f)b(j,f(j)) \leq 0 \qquad\qquad \text{for all } i \in I\ .$$

Let R(f) be the set of states which are positive recurrent under policy f, then, by assumption 2, R(f) is not empty. Let $i \in R(f)$, then $\pi_{ii}(f)>0$. Moreover, $b(j,a) \geq 0$ for all $a\in A(j)$ and $j\in I$ (see(43)). Hence it follows from (54) and (42) that

$$(55)\qquad 0=b(i,f(i)) = c(i,f(i))-g+ \sum_{j\in I} p_{ij}(f)v(j)-v(i) \quad \text{for all } i \in R(f)$$

Next it follows from lemma 2 that there is a policy $f^* \in F_{opt}$ such that each $i \in R(f)$ is positive recurrent under policy f^*. Since assumption 3 holds, it now follows from theorem 1 that $v_n^*(i)$ has a finite limit as $n \to \infty$ for each $i \in R(f)$, and so

$$(56) \qquad m(i) = M(i) \qquad \text{for all } i \in R(f) .$$

We are now ready to prove that $m(i) = M(i)$ for all $i \in I$. Using the boundedness of the sets $\{m(i)\}$ and $\{M(i)\}$, we have by (52) that

$$(57) \qquad 0 \leq M(i)-m(i) \leq \sum_{j \in I} p_{ij}(f)\{M(j)-m(j)\} \qquad \text{for all } i \in I .$$

from which it follows that

$$(58) \qquad 0 \leq M(i)-m(i) \leq \sum_{j \in I} \pi_{ij}^{(n)}(f)\{M(j)-m(j)\} \qquad \text{for all } i \in I .$$

By assumption 2, $\sum_{j \in I} \pi_{ij}(f) = 1$ for all $i \in I$, and so, $\sum_{j \in I} \pi_{ij}^{(n)}(f)$ converges uniformly with respect to n for each $i \in I$ [3, p. 37]. Using this and the boundedness of $\{M(j)-m(j)\}$, we get by a standard argument

$$(59) \qquad 0 \leq M(i)-m(i) \leq \sum_{j \in I} \pi_{ij}(f)\{M(j)-m(j)\} \qquad \text{for all } i \in I .$$

However, for each $i \in I$, $\pi_{ij}(f) = 0$ if $j \notin R(f)$, and so, by (56) and (59),

$$(60) \qquad m_i = M_i \qquad \text{for all } i \in I .$$

This proves the assertion (a) of the theorem.

(b) Let $\beta(i) = \lim_{n\to\infty} v_n^*(i)$, $i \in I$. Since $m(i)=M(i)=\beta(i)$ for all $i \in I$, we have by (45) and (46) that

(61) $$\beta(i) = \min_{a\in A(i)} \{b(i,a) + \sum_{j\in I} p_{ij}(a)\beta(j)\} \quad \text{for all } i \in I ,$$

and so, by (42),

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

Since assumption 4 is now satisfied and $\beta(i)$ is bounded, it follows from lemma 3 that, for some constant c, $\beta(i) = c$ for all $i \in I$. This ends the proof of the theorem.

References

[1] J.A. Bather, Optimal decision procedures for finite Markov chains, Technical Report No. 43, Department of Statistics, Stanford University, 1972 (to appear in Adv. App. Prob.)

[2] B.W. Brown, On the iterative method of dynamic programming on a finite space discrete time Markov process, Ann. Math. Statist. (36), 1965, 1279-1285.

[3] K.L. Chung, *Markov chains with stationary transition probabilities.* Springer Verlag, Berlin 1960.

[4] C. Derman, *Finite state Markovian decision processes.* Academic Press, New York, 1970.

[5] A. Hordijk, Over een Doeblinvoorwaarde en haar toepassing in beslissingsprocessen, Rapport BW15/72 (Prepublication), Mathematisch Centrum, Amsterdam, 1972.

[6] R.A. Howard, *Dynamic programming and Markov processes.* Technology Press, Cambridge, Mass. 1960.

[7] D.L. Iglehart, Optimality of (s,S) inventory policies in the infinite horizon dynamic inventory problem, Management Sci. 9 (1963), 259-267.

[8] D.L. Iglehart, Dynamic programming and stationary analysis of inventory problems, chap. 1 in H. Scarf, D. Gilford and M. Shelly (eds.), *Multistage inventory models and techniques*, Stanford Univ. Press. Stanford, Calif. 1963.

[9] E.L. Johnson, On (s,S) policies, Management Sci. 15 (1968), 80-101.

[10] B.A. Kalymon, Stochastic prices in a single-item inventory purchasing model, Opns. Res. 19 (1971), 1434-1458.

[11] E. Lanery, Etude asymptotique des systemes Markoviens a commande, R.I.R.O. 1, no 5 (1967), 3-56.

[12] S.M. Ross, Arbitrary state Markovian decision processes, Ann. Math. Statist. 39 (1968), 2118-2122.

[13] H. Scarf, The optimality of (S,s) policies in the dynamic inventory problem, chap. 13 in K.J. Arrow, S. Karlin and P. Suppes (eds.), *Mathematical methods in the social sciences*, Stanford Univ. Press, Stanford, Calif. 1960.

[14] P.J. Schweitzer, Perturbation theory and Markovian decision processes, MIT Operations Research Center, Technical Report No. 15, 1965.

[15] P.J. Schweitzer, Perturbation theory and undiscounted Markov renewal programming, Opns. Res. 17 (1969), 716-727.

[16] P.J. Schweitzer (personal communication), 1973.

[17] H.C. Tijms, *Analysis of (s,S) inventory models*, Mathematical Centre Tract No. 40, Mathematisch Centrum, Amsterdam, 1972.

[18] A.F. Veinott, Jr., and H.M. Wagner, Computing optimal (s,S) policies, Management Sci. 11 (1965), 525-552.