MATHEMATICS OF OPERATIONS RESEARCH
Vol. 17, No 2, May 1992
Printed in U S A

RECURRENCE CONDITIONS FOR AVERAGE AND BLACKWELL OPTIMALITY IN DENUMERABLE STATE MARKOV DECISION CHAINS*†

ROMMERT DEKKER AND ARIE HORDIJK

In a previous paper Dekker and Hordijk (1988) presented an operator theoretical approach for multichain Markov decision processes with a countable state space, compact action sets and unbounded rewards. Conditions were presented guaranteeing the existence of a Laurent series expansion for the discounted rewards, the existence of average and Blackwell optimal policies and the existence of solutions for the average and Blackwell optimality equations. While these assumptions were operator oriented and formulated as conditions for the deviation matrix, we will show in this paper that the same approach can also be carried out under recurrence conditions. These new conditions seem easier to check in general and are especially suited for applications in queueing models.

1. Introduction. In this paper we consider average and Blackwell optimality in denumerable state Markov decision chains with compact action sets. In the introduction of Dekker and Hordijk (1988) an overview of the literature with 38 references was given. Let us therefore give a brief introduction in this paper. While the theory of finite state, finite action Markov decision chains was accomplished in a relatively short time, the analysis proved to be quite difficult for the average optimality criterion in denumerable state Markov decision chains and many counterexamples for the nonexistence of average optimal policies were given. First results for the denumerable state space required a special Markov chain structure, viz. one ergodic class (the unichain case). It took some time before that case was studied in detail (see Federgruen, Hordijk and Tijms 1979). Later on, the analysis was extended to the so-called communicating case, in which a certain set of states can always be reached from any state under at least one policy (see Federgruen, Schweitzer and Tijms 1983). Although this implied a substantial relaxation, the assumptions were still not fulfilled in any finite state and finite action model. Recent research has shown that assumptions are possible which do incorporate the finite state and action model. Zijm (1985) gives an analysis for the bounded rewards case. Also more sensitive optimality criteria have attained attention. Mann (1985) gives conditions for n-discount optimality in case of bounded rewards.

Dekker and Hordijk (1988) presented a new operator theoretical approach for denumerable state Markov decision chains in which no direct assumptions were made on the Markov chain structure and which was also valid for the general finite state, finite action model. That paper only contains conditions which are related to operators, like the deviation matrix $D(f)$, and a condition on a geometric conver-

*Received March 30, 1989; revised May 17, 1990.
AMS 1980 subject classification. Primary 90C40.
IAOR 1973 subject classification. Main: Programming: Markov Decision.
OR/MS Index 1978 subject classification. Primary: 119 Dynamic Programming/Markov/Infinite State.
Key words. Denumerable Markov decision chains, average optimality, sensitive optimality criteria, optimality equation, unbounded immediate rewards, Laurent series expansion, multichain model, recurrence conditions.
†This research was partially sponsored by the Netherlands Organization for Scientific Research (NWO).

0364-765X/92/1702/0271/$01.25

gence of $P^k(f)$ towards the stationary matrix $\Pi(f)$. A relation of these conditions with the eigenvalues of $P(f)$ was given by Laserre (1988). Although the operator theoretical approach is elegant, it may be difficult to check its conditions. Recurrence conditions can be checked easier, but require a far more technical analysis as many of the preceding papers show. In this paper we present recurrence conditions which are sufficient for both average as well as Blackwell optimality. Blackwell optimality is a stronger criterion than average optimality and is also stronger than n-discount optimality. Recurrence conditions for n-discount optimality can be slightly weaker than for Blackwell optimality, but do not allow for wider problem classes.

Our recurrence conditions consist of a part requiring geometric recurrence to a finite set and a part which requires continuity of the number of classes. Most counterexamples (e.g., Hordijk and Dekker 1983) for the nonexistence of average optimal policies were based on the lack of continuity of the number of classes. As it seems difficult to check this continuity property, we use the concept of reference states to facilitate the verification. However, even in finite-state and compact-action Markov decision chains continuity of the number of ergodic classes is required (see Schweitzer 1982). In Dekker and Hordijk (1989) the results of this paper and the 1988 paper are applied and extended to semi-Markov decision chains.

Let us briefly review the literature on recurrence conditions. In Hordijk (1974) recurrence conditions have been used to establish the existence of average optimal policies in Markov decision chains. Their relation with the simultaneous Doeblin condition, Liaponuv functions and Foster criteria have been studied there. It easily follows from the analysis of Chapter 11 of Hordijk (1974) that the simultaneous Doeblin condition is equivalent to the condition,

$$\|{}_M P^n(f)\| < c\beta^n, \qquad f \in F, \quad n = 1, 2, \ldots, \tag{1.1}$$

where M is a finite set, c is a finite constant, $\beta < 1$, F is the set of all decision rules, $\|\cdot\|$ is the supremum norm and ${}_M P(f)$ is the matrix of transition probabilities under decision rule f and taboo set M, i.e.,

$${}_M P_{ij}(f) = \begin{cases} P_{ij}(f), & j \notin M, \\ 0, & j \in M. \end{cases}$$

A simple argument (cf. Lemma 5.3) gives that the relation (1.1) implies the existence of a bounded vector μ such that ${}_M P(f)$ is a contraction in a μ-weighted supremum norm with a contraction factor $\tilde{\beta} < 1$ for all decision rules, i.e.,

$$\|{}_M P(f)\|_\mu < \tilde{\beta}, \qquad f \in F. \tag{1.2}$$

In this paper we use condition (1.2) for unbounded μ. This is essential since the immediate reward vector which is allowed to be unbounded must have a finite μ-norm (condition A). For one Markov chain the condition (1.2) is closely related to a criterion of Popov (1977) for geometric ergodicity of the Markov chain. In Hordijk and Spieksma (1989) ergodicity and recurrence properties of a Markov chain and their relations to the Foster, Popov and Doeblin condition are studied. For an overview of the literature on these topics we refer to that paper.

The μ-geometric recurrence condition of this paper (Definition 3.2.1) is a stronger version of (1.2). Indeed, it is equivalent to (1.2) together with the continuity of the number of recurrent classes (see §3.2 and §5). We introduced this condition in a technical paper which appeared as Chapter 1 in Dekker (1985). In §4 the condition is

shown to hold for an exponential queueing model with controlled admission. In Schäl (1987) the relation (1.2) is verified for the $M/G/1$-queue with unknown input rate and controllable service rate. In Spieksma (1991) the μ-geometric recurrence condition is proved for two queueing models: the k competing queues and the two-center open Jackson network with control of the service rates. For Markov decision chains with the total reward criterion the following relation has been used

$$\|P(f)\|_\mu < \beta, \qquad f \in F. \tag{1.3}$$

The difference is that no taboo set M is involved, hence the matrices are assumed to be transient. For characterizations of the condition (1.3) see Van Hee and Wessels (1978) and Hordijk and Kallenberg (1984). Let us return to recurrence conditions. In Federgruen et al. (1978) and Thomas (1980) various equivalent recurrence conditions are given for the unichain Markov decision chain. In Zijm (1985) it is shown for the aperiodic multichain case that a bounded mean recurrence time to a finite set together with the continuity of the number of classes is equivalent to uniform geometric ergodicity.

Recently, it is shown in Dekker, Hordijk and Spieksma (1990) that under the aperiodicity assumption the μ-uniform geometric convergence condition of Dekker and Hordijk (1988) is equivalent to the μ-geometric recurrence condition (Definition 3.2.1), if the number of classes is finite for any deterministic policy.

The structure of this paper is as follows. In §2 we will give the model, state the problem and give the final objective of this paper. In the following section a normed linear space is introduced and a recurrence condition are given. We will show that this recurrence condition is sufficient for average and Blackwell optimality. The (mostly technical) proofs are left out of this section in order to give the reader a better overview of the method. In §4 the verification of the recurrence condition is shown in a single-server queueing system. For the application of geometric recurrence to multidimensional queues and networks of queues an equivalent but different condition is needed. In §5 we give other recurrence conditions and discuss their equivalence. Also it is shown that our μ-geometric recurrence conditions imply that the matrices of taboo probabilities are a contraction with respect to a different bounding vector. The final section contains the technical proofs.

2. The model. Consider the standard Markov decision model, consisting of the four tuple (E, A, P, r), where E is the denumerable *state space*, $A(i)$ the *set of available actions* in state $i \in E$, $P_{ij}(a)$ the *transition probability* from state i to state j under action a and $r_i(a)$ the *immediate reward* in state i when action a has been chosen. In this paper we will restrict ourselves to the class of *stationary and deterministic policies* F, with $f \in F$ if $f(i) \in A(i)$, for all $i \in E$. The extension of the results to the class of nonstationary policies follows from §5 in Dekker and Hordijk (1988). Let $P(f), r(f)$ denote the matrix of transition probabilities, vector of immediate rewards under policy f, respectively, i.e., $P_{ij}(f) = P_{ij}(f(i))$, etc. Furthermore, let $P^k(f) \equiv P(f)P^{k-1}(f)$, $k = 1, 2, \ldots$ and $P^0(f) \equiv I$, the identity matrix. The following assumption is standard in the case of a denumerable state space.

Assumption 1. (i) *$A(i)$ is a compact metric set for all $i \in E$.*

(ii) *Both $r_i(a)$ and $P_{ij}(a)$ are continuous on $A(i)$ for all $i, j \in E$.*

Throughout this paper we will assume this assumption to hold. The following infinite horizon optimality criteria will be considered. The *expected α-discounted rewards* of policy f, $v^\alpha(f)$ are defined by

$$v^\alpha(f) \equiv \sum_{k=0}^{\infty} \alpha^k P^k(f) r(f), \tag{2.1}$$

with $0 \leqslant \alpha < 1$, where the factor α is called the *discount factor*. In the sequel we also use $\rho \equiv (1-\alpha)/\alpha$ or equivalently $\alpha = 1/(1+\rho)$, where ρ is called the *interest rate*. We also write $v^\rho(f)$ for the $(1/(1+\rho))$-discounted rewards, so the symbols α and ρ denote the discounting with a factor α and $1/(1+\rho)$. A policy f_α is *α-discounted* optimal if for all $i \in E$, $f \in F$

$$v_i^\alpha(f_\alpha) \geqslant v_i^\alpha(f). \tag{2.2}$$

A second criterion is the (*long-run*) *average expected rewards* $g(f)$, defined by

$$g(f) \equiv \liminf_{N\to\infty} \frac{1}{N+1} \sum_{k=0}^{N} P^k(f) r(f). \tag{2.3}$$

A policy f_0 is said to be *average optimal*, if it maximizes $g_i(f)$ for all states $i \in E$ simultaneously. Average optimality is a rather insensitive criterion, as only the long term counts. More sensitive optimality criteria were therefore introduced, which also consider short-term rewards. Following Veinott (1969) a policy f_0 is said to be *n-discount optimal*, $n \geqslant -1$, if for all $i \in E$, $f \in F$

$$\liminf_{\rho \downarrow 0} \rho^{-n}\left[v_i^\rho(f_0) - v_i^\rho(f)\right] \geqslant 0.$$

Under mild conditions it can be shown that (-1)-discount optimality is equivalent to average optimality. Another optimality criterion is Blackwell optimality. A policy f_0 is *Blackwell optimal* if for every $i \in E$, $f \in F$ there exists a $\rho(i,f) > 0$ such that $v_i^\rho(f_0) - v_i^\rho(f) \geqslant 0$, $0 < \rho < \rho(i,f)$. If for all $i \in E$, $f \in F$ a Laurent series expansion for $v_i^\rho(f)$ exists then it can be shown that Blackwell optimality is equivalent to n-discount optimality for all n.

For each optimality criterion it is possible to define optimality equations from which optimal policies can be derived. Dekker and Hordijk (1988) studied the so-called Blackwell optimality equations, which require a Laurent series expansion for the discounted rewards. Not surprisingly, the Blackwell optimality equations contain both the average as well as the n-discount optimality equations for all n. Assuming the following Laurent series expansion for the discounted rewards (see also §3)

(2.4)

$$v^\rho(f) = (1+\rho)\left[\frac{\Pi(f)r(f)}{\rho} + \sum_{k=0}^{\infty} (-1)^k \rho^k D^{k+1}(f) r(f)\right], \qquad \rho > 0,$$

the Blackwell optimality equations in $y^{(-1)}, y^{(k)}$, $k = 0, 1, 2, \ldots$ have the following nested form

$$\max_{a \in A(i)} \left[\sum_j P_{ij}(a) y_j^{(-1)} - y_i^{(-1)}\right] = 0,$$

$$\max_{a \in A^{(-1)}(i)} \left[r_i(a) + \sum_j P_{ij}(a) y_j^{(0)} - y_i^{(0)} - y_i^{(-1)}\right] = 0,$$

where $A^{(-1)}(i)$ contains the maximizing actions of $\sum_j P_{ij}(a)y_j^{(-1)}$, and

$$\max_{a \in A^{(k-1)}(i)} \left[\sum_j P_{ij}(a) y_j^{(k)} - y_i^{(k)} - y_i^{(k-1)} \right] = 0 \quad \text{for } k = 1, 2, \ldots,$$

where $A^{(k-1)}(i)$, $k = 0, 1, \ldots$ is the subset of $A^{(k-2)}(i)$ consisting of the maximizing actions of the terms $r_i(a) + \sum_j P_{ij}(a)y_j^{(0)}$ for $k = 1$ and of $\sum_j P_{ij}(a)y_j^{(k-1)}$ for $k = 2, 3, \ldots$.

The goal of this paper is to show that under an appropriate recurrence condition (see Definition 3.2.1) solutions $y^{(-1)}, y^{(0)}, \ldots$ to these equations do exist. These solutions are elements of a normed linear space which is introduced in §3.1. Moreover, when the vectors $y^{(-1)}, y^{(0)}, \ldots, y^{(n+1)}$ are known the nonempty sets $A^{(-1)}(i), A^{(0)}(i), \ldots, A^{(n)}(i)$, $i \in E$ are available. Any decision rule $f \in F$ with $f(i) \in A^{(k)}(i)$, $i \in E$, $k \leqslant n + 1$ gives a stationary deterministic policy which is n-discount optimal. A Blackwell optimal policy is obtained by taking a decision rule f for which $f(i) \in A^{(k)}(i)$, $i \in E$ for all $k = -1, 0, 1, \ldots$. Such decision rules, called conserving decision rules, do exist (see Corollary 3.11).

3. Recurrence conditions for average and Blackwell optimality.

3.1. *Normed linear spaces.* The approach we will use to handle unbounded rewards in a denumerable state space is to introduce a weighted supremum norm $\| \cdot \|_\mu$, where μ is a vector of positive weights (also called *bounding vector*). For any vector x on E its μ-norm is defined as $\|x\|_\mu \equiv \sup_{i \in E} |x_i|/\mu_i$. For a matrix A on $E \times E$ the associated operator norm $\|A\|_\mu$ reduces to $\sup_{i \in E} \sum_j |A_{ij}|\mu_j/\mu_i$. Based on these norms we introduce normed linear spaces V^μ, M^μ of vectors, matrices on E respectively. These spaces guarantee that for any matrix $A \in M^\mu$ and vector $x \in V^\mu$ the product Ax exists and is again contained in V^μ. Our approach is to look for a vector μ such that not only $P(f)$ and $r(f)$ have finite μ-norms, but that appropriate recurrence conditions containing the vector μ hold so that all quantities of interest are properly defined.

The first quantity of interest is the matrix $\Pi(f) \equiv \lim_{\alpha \uparrow 1}(1 - \alpha)\sum_{k=0}^\infty \alpha^k P^k(f)$, which is related to the average rewards. It is called the *stationary matrix* because the following properties hold: $\Pi(f)P(f) = P(f)\Pi(f) = \Pi(f)$ (cf. Chung 1960). It can be shown that $\Pi(f)$ exists in a denumerable state space, although it may be defective. A second quantity of interest, which is related to more sensitive criteria, is the *deviation matrix* $D(f)$, defined as

$$D(f) \equiv \lim_{\alpha \uparrow 1} \sum_{k=0}^{\infty} \alpha^k \left[P^k(f) - \Pi(f) \right] \tag{3.1}$$

provided the limit exists. The importance of $D(f)$ lies in its relation with the Laurent expansion of the discounted rewards. To state this precisely we need the following condition

Condition A. *For all $f \in F$ the following assertions hold*:

(i) $\|r(f)\|_\mu < \infty$,

(ii) $\sup_{0 < \alpha < 1}(1 - \alpha)\|\sum_{k=0}^\infty \alpha^k P^k(f)\|_\mu < \infty$,

(iii) $D(f)$ *exists*, $\|D(f)\|_\mu < \infty$ *and the following relations are valid*

$$D(f)[I - P(f)] = [I - P(f)]D(f) = I - \Pi(f),$$

$$\Pi(f)D(f) = D(f)\Pi(f) = 0.$$

In Dekker and Hordijk (1988, Theorem 4.1) the following result is shown.

THEOREM 3.1. *If Condition* A *holds then* $v^\rho(f)$ *has the Laurent series expansion* (2.4).

3.2. *Recurrence conditions.* In this section we will introduce recurrence conditions for average as well as Blackwell optimality. First we recapitulate briefly the Markov chain concepts necessary for our analysis (cf. Chung 1960). We call a set of essential states a *class* if from every state within the class every other state in that class can be reached and no other states. A Markov chain can be divided into classes of essential states and inessential states which do not belong to a class. In this paper we assume strong recurrence to a set of reference states (Assumption 2(ii)). It follows from Lemma 3.1 and Theorem 3.5(iii) that the classes consist of positive recurrent states and hence the inessential states are the transient states. For a subset B of E we denote by $F_{iB}(f)$ the probability that set B is eventually reached from state i under policy f. Let the matrix ${}_BP(f)$ with elements ${}_BP_{ij}(f)$, $i, j \in E$ be defined as

$$ {}_BP_{ij}(f) \equiv \begin{cases} P_{ij}(f), & j \notin B, \\ 0, & j \in B. \end{cases} $$

$({}_BP(f))^k$ is abbreviated to ${}_BP^k(f)$. Following Zijm (1985) we will call a set $B(f) \subset E$ a *set of reference states for policy* f if it contains exactly one state of each class under policy f and no other states.

The recurrence condition we consider is called *μ-geometric recurrence* to a set of reference states (abbreviated to μ-GRRS).

DEFINITION 3.2.1. A set of Markov chains $P(f)$, $f \in F$ satisfies condition μ-GRRS if there exist a finite set $M \subset E$, a constant c and a $\beta < 1$ such that for each $f \in F$ there exists a set of reference states $B(f) \subset M$ with

$$ \left\| {}_{B(f)}P^n(f) \right\|_\mu \leqslant c\beta^n, \qquad n = 1, 2, \ldots . $$

If the μ-geometric recurrence property holds for a specific set M and constants c and β, it will be denoted by μ-GRRS(M, c, β). In §5 we will discuss equivalent conditions some of which are easier to verify. We will now formulate the main assumption under which we will develop the first part of our results.

Assumption 2. *There exist a vector μ of positive weights ($\geqslant 1$), a finite set M and finite constants c_1 and $\beta < 1$ such that*

(i) $\|r(f)\|_\mu < \infty$, $f \in F$,

(ii) *condition μ-GRRS(M, c_1, β) holds.*

Notice that Assumption 2 implies that $\|P(f)\|_\mu < \infty$, $f \in F$. Although condition μ-GRRS is formulated as a condition uniform in f, the uniformness is only required in the next section.

Throughout the rest of this paper we denote by $B(f)$ the set of reference states mentioned in Assumption 2. Consequences of this assumption are stated in the following lemmas.

LEMMA 3.1. *For each policy $f \in F$ and state $i \in E$ we have $F_{iB(f)}(f) = 1$.*

LEMMA 3.2. $\sup_{k=1,2,\ldots} \sup_{f \in F} \|P^k(f)\|_\mu < \infty$.

Note that Lemma 3.2 implies Condition A(ii). Remark further that in Assumption 2 the conditions on the Markov chain $P(f)$ and on the reward vector $r(f)$ are separated. This implies that our analysis remains the same if we replace $r(f)$ by any other μ-bounded reward vector r. We will make use of this fact by considering both average rewards $g(f, r)$ and discounted rewards $v^\rho(f, r)$ where the μ-bounded

vector r represents the immediate rewards. In the sequel we will establish properties for the matrices $\Pi(f)$ and $D(f)$ from properties for $g(f,r)$ and $v^{\rho}(f,r)$ by taking suitable reward vectors r. Let $\delta(j)$ be the jth unit vector on E, i.e., only the jth element of $\delta(j)$ is nonzero and equal to 1 and let e denote the vector with all elements equal to one. Note that both $\delta(j)$ and e are μ-bounded vectors.

We will first show that Assumption 2 is sufficient to establish the Laurent series expansion (2.4). We start with developing a partial Laurent series expansion by using a technique from Hordijk and Sladky (1977). From this partial Laurent series explicit expressions for $\Pi(f)$ and $D(f)$ can be obtained, with which the relations of Condition A(iii) can be proven and the total Laurent series expansion can be established.

As we will consider one policy f only in the remainder of this section we denote its classes by $C_1,\dots,C_{\nu(f)}$, where $\nu(f)$ is the number of classes under policy f. Let b_j be the reference state contained in class C_j, i.e., $b_j = C_j \cap B(f)$ and let $T(f)$ denote the set of transient states. The following lemma gives explicit expressions for what later on will turn out to be terms of the Laurent series expansion.

LEMMA 3.3. *For any vector $y \in V^{\mu}$ the following set of equations in u, v*

$$\begin{cases} u = P(f)u, \\ u + v = P(f)v + y, \\ v = 0 \quad \text{on } B(f), \end{cases} \tag{3.2}$$

has unique solutions $u(f,y), v(f,y)$ in V^{μ} given by

$$\begin{cases} u_i(f,y) = \sum_{k=0}^{\infty} \left(P_{B(f)}^k(f)y\right)_{b_j} \Big/ \sum_{k=0}^{\infty} \left(P_{B(f)}^k(f)e\right)_{b_j}, & i \in C_j, \\ u_i(f,y) = \sum_{j=1}^{\nu(f)} F_{ib_j}(f)u_{b_j}(f,y), & i \in T(f), \\ v(f,y) = \sum_{k=0}^{\infty} P_{B(f)}^k(f)[y - u(f,y)]. \end{cases} \tag{3.3}$$

Moreover, the following decomposition of the solutions is valid

$$u(f,y) = \sum_{i \in E} u(f,\delta(i))y_i, \qquad v(f,y) = \sum_{i \in E} v(f,\delta(i))y_i.$$

Note that by Assumption 2 all expressions are finite, that the sums are absolutely convergent and that $u(f,r)$ is constant on each class. To establish a partial Laurent series expansion, we first consider the discounted rewards for a finite horizon, $v^{\alpha,T}(f,r) \equiv \sum_{k=0}^{T-1} \alpha^k P^k(f)r$. Denote the solutions u, v from Lemma 3.3 with $y = r$ by $u^{(-1)}(f,r)$ and $v^{(0)}(f,r)$. We now replace r in the expression for $v^{\alpha,T}(f,r)$ by $u^{(-1)}(f,r) + v^{(0)}(f,r) - P(f)v^{(0)}(f,r)$ and rearrange terms. We again apply Lemma 3.3 now with $y = v^{(0)}(f,r)$ and denote the solutions by $u^{(0)}(f,r)$ and $v^{(1)}(f,r)$. It turns out (see the proof of Theorem 3.4) that by Assumption 2 the limit for $T \to \infty$ can be taken which yields the following theorem.

THEOREM 3.4. *Under Assumption 2 there exists the following expansion*

$$v^{\alpha}(f,r) = (1+\rho)\left[\frac{u^{(-1)}(f,r)}{\rho} + v^{(0)}(f,r) - u^{(0)}(f,r) - \rho v^{(1)}(f,r)\right] + \rho^2 \sum_{k=0}^{\infty} \alpha^k P^k(f)v^{(1)}(f,r). \tag{3.4}$$

From this expansion we will show the properties of the matrices $\Pi(f)$ and $D(f)$. Let $s^\alpha(f,r) \equiv (1-\alpha)v^\alpha(f,r)$ and $w^\alpha(f,r) \equiv v^\alpha(f,r) - g(f,r)/(1-\alpha)$. Lemma 3.3 and Theorem 3.4 lead us to the main theorem of this section.

THEOREM 3.5. *Under Assumption* 2 *we have*
(i) $g(f,r) = \lim_{\alpha\uparrow 1} s^\alpha(f,r) = u^{(-1)}(f,r) = \Pi(f)r$, $f \in F$, $r \in V^\mu$.
(ii) $D_{ij}(f)$ *exists for all* $i, j \in E$, $f \in F$ *and*

$$D(f)r = \lim_{\alpha\uparrow 1} w^\alpha(f,r) = v^{(0)}(f,r) - u^{(0)}(f,r),\ f \in F,\ r \in V^\mu.$$

(iii)

$$\lim_{\alpha\uparrow 1} \sup_{f\in F} \sup_{\|r\|_\mu \le 1} \|s^\alpha(f,r) - \Pi(f)r\|_\mu$$

$$= \lim_{\alpha\uparrow 1} \sup_{f\in F} \sup_{\|r\|_\mu \le 1} \|w^\alpha(f,r) - D(f)r\|_\mu = 0.$$

(iv) $\sup_{f\in F} \|D(f)\|_\mu < \infty$.

REMARK. In Theorem 3.5 we gave explicit expressions for $\Pi_{ij}(f)$ and $D_{ij}(f)$ via the solutions of Lemma 3.3. These solutions do not depend on the chosen set of reference states as $\Pi(f)$ and $D(f)$ are defined via Abelian limits. In fact any set of reference states for which the sums in (3.3) are well defined give the same solution. It will be shown that Condition A(iii) easily follows from part (iii), hence we have the following corollary.

COROLLARY 3.6. *Under Assumption* 2 *the matrix* $D(f)$ *defined in* (3.1) *exists and* $v^\rho(f)$ *has the Laurent series expansion given in* (2.4).

3.3. *Optimality*. Where we showed the existence of a Laurent series expansion for a single policy in the previous section, we will deal with all stationary and deterministic policies in this section and derive optimality results. The main concept we will use to establish optimality in a denumerable state space and compact action set is continuity. Recall that the policy space F is defined as $\prod_{i=1}^\infty A(i)$, the product space of the sets of available actions in each state and that F is endowed with the product topology. Hence a sequence of policies converges, say $f^{(n)} \to f^{(0)}$ if $f^{(n)}(i) \to f^{(0)}(i)$ on $A(i)$, $i \in E$. If both E and $A(i)$ for all $i \in E$ are finite then F is a finite set and any function of F is trivially continuous. For a finite state space and compact action set Schweitzer (1982) showed the existence of an optimal policy. Besides continuity of both the immediate rewards and the transition probabilities, his assumptions also required the continuity of the number of classes $\nu(f)$. However, this approach cannot directly be used in a denumerable state space as the standard continuity concept is not sufficient: e.g., continuity of $P_{ij}(f)$ and $r_i(f)$ for all $i, j \in E$ does not imply continuity of its product $\Sigma_j P_{ij}(f)r_j(f)$. Therefore Dekker and Hordijk (1988) introduced the concept of μ-continuity which is related to the μ-norm used. It is defined as

DEFINITION 3.3. A matrix function $A(f) \in M^\mu$ is μ-continuous on F if for every state $i \in E$ and sequence $f^{(n)} \to f^{(0)}$ in F we have

$$\lim_{f^{(n)}\to f^{(0)}} \sum_j \left| A_{ij}(f^{(n)}) - A_{ij}(f^{(0)}) \right| \mu_j = 0.$$

The following lemmas can be proven by standard analysis. Detailed proofs are given in Dekker (1985).

LEMMA 3.7. *For any matrix function* $A(f) \in M^{\mu}$ *assertions* (i), (ii) *and* (iii) *are equivalent*:

(i) $A(f)$ *is* μ*-continuous on* F.

(ii) *Both* $A(f)$ *and* $|A(f)|\mu$ *are pointwise continuous on* F.

(iii) *For any sequence* $x^{(n)} \to x^{(0)}$, *pointwise converging on* E *with*

$$\sup_{n=0,1,2,\ldots} \|x^{(n)}\|_{\mu} < \infty,$$

it holds that $A(f^{(n)})x^{(n)} \to A(f^{(0)})x^{(0)}$ *pointwise on* E *for every sequence* $f^{(n)} \to f^{(0)}$ *on* F.

LEMMA 3.8. *If both* $A(f)$ *and* $B(f) \in M^{\mu}$ *are* μ*-continuous on* F, *then*

(i) $A(f) + B(f)$ *is also* μ*-continuous on* F.

(ii) *If* $\sup_{f \in F} \|B(f)\|_{\mu} < \infty$, *then* $A(f)B(f)$ *is also* μ*-continuous on* F.

We will now formulate the assumption which we require to establish continuity of the various quantities of interest.

Assumption 3. (i) $P(f)\mu$ *is pointwise continuous on* F.

(ii) *There exists some constant* c_0 *such that* $\|r(f)\|_{\mu} < c_0$, $f \in F$.

Note that by Lemma 3.7 the μ-continuity of $P(f)$ follows from Assumption 3(i). Remark further that by Assumption 3(i), Lemmas 3.2 and 3.8 the same holds for $P^k(f)$, $k = 2, 3, \ldots$. It can be shown (cf. Dekker 1985) that Assumptions 2 and 3 incorporate continuity of the number of classes and thus include the model of Schweitzer (1982).

Our optimality results will be based on Theorem 4.7 of Dekker and Hordijk (1988). In terms of this paper it reads:

THEOREM 3.9. *If Condition* A(iii) *and Assumption* 3 *hold and if* $D(f)$ *is* μ*-continuous and has a bounded* μ*-norm then there exist unique solutions to the Blackwell optimality equations.*

The boundedness of $\|D(f)\|_{\mu}$ in f is already shown in Theorem 3.5. The μ-continuity of $D(f)$ is a consequence of the uniform convergence in μ-norm of the defining series of $\Pi(f)$ and $D(f)$.

THEOREM 3.10. (i) $\Pi(f)$ *is* μ*-continuous.*

(ii) $D(f)$ *is* μ*-continuous.*

The main theorem of this paper is the following corollary.

COROLLARY 3.11. *Assumptions* 1, 2 *and* 3 *are sufficient to establish the Blackwell optimality equations, with the following consequences*

(i) *there exist unique* μ*-bounded solutions* $y^{(-1)}, y^{(0)}, \ldots$ *to them.*

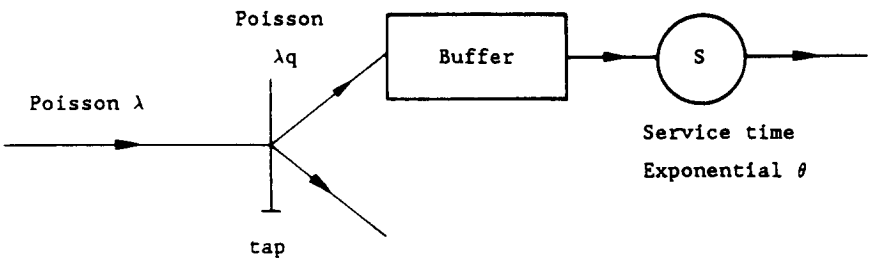
(ii) *For* $n = -1, 0, 1, \ldots$, *any policy* f_0 *with* $f_0(i) \in A^{(n+1)}(i)$ *is n-discount optimal.*

(iii) *The intersections* $A^{(\infty)}(i) \equiv \bigcap_{k=-1}^{\infty} A^{(k)}(i)$, $i \in E$ *are nonempty. Any conserving decision rule* f_0, *i.e.*, $f_0(i) \in A^{(\infty)}(i)$, $i \in E$ *gives a stationary deterministic Blackwell optimal policy.*

We would like to remark that not always all parts of the Blackwell optimality equations have to be checked. If after two equations for each state i the corresponding action set $A^{(0)}(i)$ is reduced to one action only, then all other equations are automatically satisfied. Hence if there exists a unique 0-discount optimal policy, it is automatically Blackwell optimal.

4. Admission control of a single-server queueing system. In this section we demonstrate the ease of checking the recurrence condition μ-GRRS for a simple queueing system. In its simplest form it does not exhibit a multichain structure, but extensions of the model can be given which do.

We are concerned with a waiting line with controllable input, as is shown in the following figure.



Assume that the arrival process is a Poisson process with parameter λ. The server can control the number of arrivals by a tap and allow only a fraction q of the arrivals to enter the system, where $0 \leqslant a \leqslant q \leqslant b \leqslant 1$, for some fixed numbers a and b. When $a > 0$, the arrival stream can never be stopped completely and any number of customers in the buffer is possible. Consequently we have a denumerable state space. The server observes the system at discrete time points $t = 0, 1, \ldots$. If at time t the server finds a nonempty buffer, he starts serving until the customer is finished or until $t + 1$, whichever comes first. The processing time S is exponentially distributed with parameter $\theta > 0$.

It would be more natural to formulate this waiting line model as a continuous time Markov decision chain. However, if we take a small time scale, the Markov decision chain is a good approximation. There is also another reason to use a small time scale. In the proof of Lemma 4.1 below we need that $\lambda b + e^{-\theta} < 1$. Assuming that the traffic intensity $\lambda/\theta < 1$, it is easily verified that for fixed traffic intensity but λ and θ small enough, that means for the time scale small enough, the inequality $\lambda b + e^{-\theta} < 1$ is always satisfied. Hence without loss of generality we assume that this inequality holds.

The transition probabilities for this system are

$$P_{ij}(q) = \begin{cases} (\lambda q)^j e^{-\lambda q}/j!, & i = 0, \\ (1 - e^{-\theta}) e^{-\lambda q}, & i \neq 0,\ j = i - 1, \\ (1 - e^{-\theta}) \dfrac{(\lambda q)^{j-i+1} e^{-\lambda q}}{(j - i + 1)!} + e^{-\theta} \dfrac{(\lambda q)^{j-i} e^{-\lambda q}}{(j-1)!}, & i \neq 0,\ j > i - 1. \end{cases}$$

It is clear that $P_{ij}(q)$ is a continuous function of q. We assume that the rewards can be written as

$$r_i(q) = \begin{cases} r - r_q - c(i-1), & i > 0, \\ -r_q, & i = 0, \end{cases}$$

where r_q denotes the cost of using control q, r is the price a customer pays for being served during one period and c is the cost of having a customer wait during one period. We assume that r_q depends continuously on q.

In search of a suitable bounding vector we try $\mu_i = x^i$, $i \in E$ with $x > 1$ and find

$$\frac{1}{\mu_i}\sum_j P_{ij}(q)\mu_j = \begin{cases} e^{(x-1)\lambda q}, & i = 0, \\ \frac{1}{x}e^{(x-1)\lambda q} + \left(1 - \frac{1}{x}\right)e^{(x-1)\lambda q - \theta}, & i > 0. \end{cases}$$

Hence for any $x > 1$, $\mu_i = x^i$, $i \in E$ is a bounding vector for the rewards and satisfies Assumptions 2(ii) and 3. We also see that $P(q)\mu$ is pointwise continuous in q. Moreover, with this μ vector we can also show μ-geometric recurrence.

LEMMA 4.1. *The Markov chains in our example are μ-geometric recurrent with M being state 0, $\mu_i = x^i$, $i \in E$, for some $x > 1$ and some $\beta < 1$.*

Accordingly it is clear that Assumptions 2 and 3 are fulfilled and that Blackwell optimality equations can be established in this case. It is plausible that the structure of a Blackwell optimal policy is of a bang-bang type. However, as far as we know, a proof of that is yet to be found in literature.

5. Equivalent recurrence conditions. In §4 the μ-geometric recurrence was verified for the admission control of a single-server queueing system. In the papers of Hordijk and Spieksma (1989), Spieksma (1991) multidimensional queues are studied. For various versions of the k-competing queues model and for networks of queues with two nodes the μ-geometric recurrence can be established. Whereas in the first paper one Markov chain is analyzed, the second paper includes the control of service rates. The extension to open exponential networks of queues with more than two nodes and the existence of average and Blackwell optimal policies for the control of these networks is as far as we know still an open problem.

It turned out that for the above-mentioned queueing systems it is easier to verify a different recurrence condition, which we called μ-bounded recurrence. In this section we introduce this condition together with μ-uniform recurrence and we show that both are equivalent to μ-geometric recurrence.

DEFINITION 5.1. Let $M \subset E$ be a finite set; a set of Markov chains $P(f)$, $f \in F$, is said to satisfy condition

(i) μ-uniform recurrence (μ-URRS(M)), if there exist an $n_0 > 0$, a $c_2 > 0$ and a $\beta < 1$ such that for all $f \in F$ there exists a set of reference states $B(f) \subset M$ with

$$\|{}_{B(f)}P^{n_0}(f)\|_\mu \leqslant \beta \quad \text{and} \quad \|P(f)\|_\mu < c_2.$$

(ii) μ-bounded recurrence (μ-BRRS(M)), if there exists a constant c_3 such that for all $f \in F$ there exists a set of reference states $B(f) \subset M$ with

$$\left\| \sum_{n=0}^{\infty} {}_{B(f)}P^n(f) \right\|_\mu < c_3.$$

THEOREM 5.2. *The conditions μ-URRS(M), μ-BRRS(M) and μ-GRRS(M) are equivalent.*

The recurrence conditions of this paper have all $B(f)$ as set of taboo states, i.e., matrices ${}_{B(f)}P(f)$ are considered. Apparently weaker conditions are obtained if we use M as taboo states. We say that the set $P(f)$, $f \in F$ satisfies condition μ-GR(M)

if for a constant c and a $\beta < 1$,

$$\|_M P^n(f)\|_\mu < c\beta^n, \qquad f \in F, \quad n = 1, 2, \ldots . \tag{5.1}$$

Note that in the notation the *RS* of reference state are omitted. Similarly conditions μ-*UR*(M) and μ-*BR*(M) can be introduced. Since $B(f) \subset M$ for all $f \in F$, it is obvious that the conditions of the Definitions 3.2.1 and 5.1 are stronger. In Dekker, Hordijk and Spieksma (1990) it will be shown that under Assumptions 1 and 3, μ-*GRRS*(M) is equivalent to μ-*GR*(M) together with the continuity of the number of classes. Similar equivalence relations hold for the other criteria. If $c < 1$ in the relation (5.1) then all matrices ${}_M P(f)$ are contractions in μ-norm with the same contraction modulus $\beta < 1$. With a standard argument (cf. Van Nunen and Wessels 1977, Lemma 5.4) one can show that it is possible to find a different bounding vector $\tilde{\mu}$ for which this holds.

LEMMA 5.3. *Under the relation* (5.1) *there are a* $\tilde{\beta} < 1$ *and a bounding vector* $\tilde{\mu}$ *such that* $\|_M P(f)\|_{\tilde{\mu}} < \tilde{\beta}$, $f \in F$.

6. Proofs. In this section we will provide the proofs.

PROOF OF LEMMA 3.1. Suppose $F_{iB(f)}(f) < 1$ for some state $i \in E$ and policy $f \in F$. Let $F^{(n)}_{iB(f)}(f)$ denote the probability of reaching set $B(f)$ from state i after exactly n time units. Note that $F_{iB(f)}(f) = \sum_{n=1}^{\infty} F^{(n)}_{iB(f)}(f)$. Since

$$\sum_j \left({}_{B(f)}P^n(f)\right)_{ij} = 1 - \sum_{k=1}^{n-1} F^{(k)}_{iB(f)}(f),$$

we have $\liminf_{n\to\infty} \sum_j ({}_{B(f)}P^n(f))_{ij} > 0$ and, accordingly, $\liminf_{n\to\infty} ({}_{B(f)}P^n(f)\mu)_i > 0$, which contradicts condition μ-*GRRS*(M).

PROOF OF LEMMA 3.2. Consider a policy f and abbreviate the set of reference states $B(f)$ to B. By last exit decomposition with respect to B we have for $k \geqslant 1$

$$\begin{aligned}
\left(P^k(f)\mu\right)_j &= \left({}_B P^k(f)\mu\right)_j + \sum_{m=0}^{k-1} \sum_{i\in B} \left(P^{k-m}(f)\right)_{ji} \left({}_B P^m(f)\mu\right)_i \\
&\leqslant \left({}_B P^k(f)\mu\right)_j + \sum_{m=0}^{k-1} \sum_{i\in B} \left({}_B P^m(f)\mu\right)_i \\
&\leqslant c_1\beta^k\mu_j + \sum_{i\in B} \frac{c_1}{1-\beta}\mu_i \\
&\leqslant \frac{c_1}{1-\beta}\left\{1 + \sum_{i\in B} \frac{\mu_i}{\mu_j}\right\}\mu_j, \qquad j \in E.
\end{aligned}$$

Since $B \subset M$ and M is a finite set and furthermore, $\mu \geqslant 1$, we have $\sup_{j\in E}\{1 + \sum_{i\in M}\mu_i/\mu_j\} < \infty$. The lemma now follows since the upper bound does not depend on $f \in F$.

PROOF OF LEMMA 3.3. From (3.3) we see that $u(f, y)$ is constant on each class C_j, $u(f, y) = u_{b_j}(f, y)e$ on C_j, hence $P(f)u(f, y) = u(f, y)$ on each C_j. Since $\sum_i P_{li}(f)F_{ib_j}(f) = F_{lb_j}(f)$, $l \in E$, we also have $u(f, y) = P(f)u(f, y)$ on E. On C_j we

have from $u(f, y) = u_{b_j}(f, y)e$

$$v(f, y) = \sum_{k=0}^{\infty} {}_{B(f)}P^k(f)y - u_{b_j}(f, y) \sum_{k=0}^{\infty} {}_{B(f)}P^k(f)e.$$

Hence $v_{b_j}(f, y) = 0$. Since $v(f, y) = {}_{B(f)}P(f)v(f, y) + (y - u(f, y))$, the second equation of (3.2) is also verified. The μ-boundedness of $u(f, y)$ and $v(f, y)$ follows from condition μ-*GRRS*(M) and the fact that the denominator $(\sum_{k=0}^{\infty} {}_{B(f)}P^k(f)e)_j \geqslant 1$, $j \in E$.

Let u, v be any solution to (3.2) and let $\tilde{u} = u - \sum_{j=1}^{\nu(f)} F_{ib_j}(f)u_{b_j}$, then $\tilde{u} = P(f)\tilde{u}$. Since $\tilde{u} = 0$ on $B(f)$ we also have $\tilde{u} = {}_{B(f)}P(f)\tilde{u}$ and $\tilde{u} = {}_{B(f)}P^n(f)\tilde{u}$ for every n. However, this implies that $\|\tilde{u}\|_\mu \leqslant \|{}_{B(f)}P^n(f)\|_\mu \|\tilde{u}\|_\mu$ hence $\tilde{u} = 0$ by condition μ-*GRRS*(M). Accordingly, $u = \sum_{j=1}^{\nu(f)} F_{ib_j}(f)u_{b_j}$, implying that u is constant on each class C_j. From the last two equations of (3.2) it follows that $v = {}_{B(f)}P(f)v + (y - u)$. Iterating this n times yields $v = {}_{B(f)}P^{n+1}(f)v + \sum_{k=0}^{n} {}_{B(f)}P^k(f)(y - u)$. Since $\lim_{n\to\infty} {}_{B(f)}P^n(f)v = 0$ by condition μ-*GRRS*(M) and $\sum_{k=0}^{\infty} {}_{B(f)}P^k(f)(y - u)$ exists, v is equal to the latter. Since u is constant on each class, say $u = u_{b_j}e$ on class C_j, we have

$$v = \sum_{k=0}^{\infty} {}_{B(f)}P^k(f)y - u_{b_j} \sum_{k=0}^{\infty} {}_{B(f)}P^k(f)e \quad \text{on } C_j.$$

From equation $v = 0$ on $B(f)$ it now follows that the solutions u, v to (3.2) are given by (3.3). The decomposition of the solutions follows directly from their representation (3.3) and the dominated convergence theorem.

PROOF OF THEOREM 3.4. Inserting $r = u^{(-1)}(f, r) + v^{(0)}(f, r) - P(f)v^{(0)}(f, r)$ in the expression for $v^{\alpha,T}(f, r)$ yields

$$v^{\alpha,T}(f, r) = \sum_{k=0}^{T-1} \alpha^k P^k(f)r = \sum_{k=0}^{T-1} \alpha^k P^k(f)[u^{(-1)}(f, r) + v^{(0)}(f, r) - P(f)v^{(0)}(f, r)]$$

$$= u^{(-1)}(f, r)\left(\frac{1 - \alpha^T}{1 - \alpha}\right) + \left(1 - \frac{1}{\alpha}\right) \sum_{k=0}^{T} \alpha^k P^k(f)v^{(0)}(f, r)$$

$$+ \frac{1}{\alpha} P^0(f)v^{(0)}(f, r) - \alpha^T P^T(f)v^{(0)}(f, r).$$

Substituting $v^{(0)}(f, r) = u^{(0)}(f, r) + v^{(1)}(f, r) - P(f)v^{(1)}(f, r)$ in the second term yields

$$v^{\alpha,T}(f, r) = u^{(-1)}(f, r)\left(\frac{1 - \alpha^T}{1 - \alpha}\right) + \frac{1}{\alpha} v^{(0)}(f, r) - \alpha^T P^T(f)v^{(0)}(f, r)$$

$$+ \left(1 - \frac{1}{\alpha}\right)\left[u^{(0)}(f, r)\left(\frac{1 - \alpha^{T+1}}{1 - \alpha}\right) + \frac{1}{\alpha} v^{(1)}(f, r) - \alpha^{T+1} P^{T+1}(f)v^{(1)}(f, r)\right]$$

$$+ \left(1 - \frac{1}{\alpha}\right)^2 \sum_{k=0}^{T+1} \alpha^k P^k(f)v^{(1)}(f, r).$$

Since $v^{(1)}(f, r)$ is μ-bounded, $\sum_{k=0}^{\infty} \alpha^k P^k(f)v^{(1)}(f, r)$ exists for each $\alpha > 0$ by the μ-boundedness of $\|P^k(f)\|_\mu$, cf. Lemma 3.2. This lemma also implies that both

$\lim_{T\to\infty} \alpha^T P^T(f)v^{(0)}(f,r) = 0$ and $\lim_{T\to\infty} \alpha^{T+1}P^{T+1}(f)v^{(1)}(f,r) = 0$. Consequently, if we let $T \to \infty$ in the last expression for $v^{\alpha,T}(f,r)$ we obtain

$$v^{\alpha}(f,r) = u^{(-1)}(f,r)\left(\frac{1}{1-\alpha}\right) + \frac{1}{\alpha}v^{(0)}(f,r) + \frac{1}{\alpha}\left(1-\frac{1}{\alpha}\right)v^{(1)}(f,r)$$

$$+\left(1-\frac{1}{\alpha}\right)\frac{u^{(0)}(f,r)}{1-\alpha} + \left(1-\frac{1}{\alpha}\right)^2 \sum_{k=0}^{\infty} \alpha^k P^k(f)v^{(1)}(f,r).$$

Rewriting α into $1/(1+\rho)$ yields (3.4).

PROOF OF THEOREM 3.5. From Lemma 3.3 and Theorem 3.4 it follows that $\Pi_{ij}(f) = \lim_{\alpha\uparrow 1} s_i^{\alpha}(f,\delta(j)) = u_i^{(-1)}(f,\delta(j))$ and that, for $r \in V^{\mu}$, $\sum_j \Pi_{ij}(f)r_j = \sum_j u_i^{(-1)}(f,\delta(j))r_j = u_i^{(-1)}(f,r)$. Remark that by an equivalence between Abelian and Tauberian limits (cf. Titchmarsh 1939, pp. 224–229) we have an alternative expression for $\Pi_{ij}(f)$, i.e.

$$\Pi_{ij}(f) = \lim_{N\to\infty} \frac{1}{N+1} \sum_{k=0}^{N} P_{ij}^k(f) = g_i(f,\delta(j))$$

and from the same equivalence $u_i^{(-1)}(f,\mu) = g_i(f,\mu)$. The first equality now follows from dominated convergence.

To prove the convergence of the defining limits for $D_{ij}(f)$ we rewrite the expression for $w^{\alpha}(f,r)$. By absolute convergence, we have for $\alpha < 1$,

$$w^{\alpha}(f,r) = \sum_{k=0}^{\infty} \alpha^k P^k(f)r - \frac{1}{1-\alpha}\Pi(f)r = \sum_{k=0}^{\infty} \alpha^k \left[P^k(f) - \Pi(f)\right]r.$$

On the other hand, it follows from (3.4) and Lemma 3.2 that

$$\lim_{\alpha\uparrow 1} w^{\alpha}(f,r) = v^{(0)}(f,r) - u^{(0)}(f,r).$$

Hence by substituting $r = \delta(j)$ we find that the matrix $D(f)$ is well defined and the equalities of the second assertion follow now in the same way as for the first assertion. From Assumption 2 and Lemma 3.3 it follows that

$$\|u(f,y)\|_{\mu} \leqslant \frac{c_1}{1-\beta}\|y\|_{\mu}\left(\sup_{i\in M} \mu_i\right) \quad \text{and} \tag{6.1}$$

$$\|v(f,y)\|_{\mu} \leqslant \frac{c_1}{1-\beta}\left(\|y\|_{\mu} + \|u(f,y)\|_{\mu}\right). \tag{6.2}$$

Hence, $\sup_{f\in F}\sup_{\|r\|_{\mu}\leqslant 1}\|v^{(0)}(f,r)\|_{\mu} < \infty$ and the same can be shown for $u^{(0)}(f,r)$ and $v^{(1)}(f,r)$. The third assertion now follows from Lemma 3.2 and this uniform boundedness in μ-norm. The last assertion is a consequence of part (ii) and the inequalities (6.1) and (6.2).

PROOF OF COROLLARY 3.6. Let $W^{\alpha}(f)$ be the matrix with $w_i^{\alpha}(f,\delta(j))$ as (i,j)th element. Since

$$\lim_{\alpha\uparrow 1}\sup_{f\in F}\|W^{\alpha}(f) - D(f)\|_{\mu} = \lim_{\alpha\uparrow 1}\sup_{f\in F}\sup_{\|r\|_{\mu}\leqslant 1}\|w^{\alpha}(f,r) - D(f)r\|_{\mu} = 0$$

by Theorem 3.5, $W^{\alpha}(f)$ converges in μ-norm to $D(f)$, uniformly in f. The equalities in Condition A(iii) can now be proven similarly to the finite state case. For example, let α be sufficiently close to 1 that $\|W^{\alpha}(f) - D(f)\|_{\mu} < \epsilon$. Hence

$$\|\Pi(f)D(f)\|_{\mu} = \|\Pi(f)[D(f) - W^{\alpha}(f) + W^{\alpha}(f)]\|_{\mu}$$

$$\leqslant \epsilon \cdot \|\Pi(f)\|_{\mu} \quad \text{for any } \epsilon > 0,$$

since $\Pi(f)W^{\alpha}(f) = 0$.

As the other parts of Condition A already have been established, the assertion follows from Theorem 3.1.

PROOF OF THEOREM 3.10. The μ-continuity of $D(f)$ is induced by the μ-continuity of $W^{\alpha}(f)$, since by Theorem 3.5(iii) the convergence is uniform in f. The μ-continuity of $W^{\alpha}(f)$ itself can be derived from the μ-continuity of both $P^{k}(f)$ and of $\Pi(f)$. As the μ-continuity of $P^{k}(f)$ is a direct consequence of Assumptions 2(ii), 3(i) and Lemma 3.8, what remains to show is the μ-continuity of $\Pi(f)$. Again this follows from the μ-continuity of $(1-\alpha)\sum_{k=0}^{\infty}\alpha^{k}P^{k}(f)$ and Theorem 3.5(iii). However, we prefer to show this through the explicit expression for $\Pi(f)r$ given in Lemma 3.3 (cf. Theorem 3.5(i)). Indeed, if we show the pointwise continuity of $\Pi(f)r$ for any $r \in V^{\mu}$ (including therefore $r = \mu$ and $r = \delta(j)$, $j = 1, \ldots$) we have established the μ-continuity of $\Pi(f)$ by Lemma 3.7, since $\Pi(f) \geqslant 0$.

Now suppose $\{f^{(m)}\}$, $m = 1, 2, \ldots$ is a sequence of policies converging to policy f for some $f \in F$. Let $B(f^{(m)})$ be the set of reference states for policy $f^{(m)}$ mentioned in Assumption 2. Since $B(f^{(m)}) \subset M$ for all m and M is a finite set, there exist a subsequence $\{f^{(m_k)}\}$, $k = 1, 2, \ldots$ and a set $B \subset M$ such that $B(f^{(m_k)}) = B$, $k = 1, 2, \ldots$. The rest of the proof now consists of the following parts. First we show that B is also a set of reference states for policy f for which the recurrence part of the condition μ-*GRRS*(M) holds. Therefore we may use B as set of reference states in (3.3) (see the remark at the end of §3.2).

Secondly we show that, for each state i, $(\Pi(f^{(m_k)})r)_i$ converges and that the limit is equal to $(\Pi(f)r)_i$. As this reasoning can also be applied for any subsequence $f^{(m_l)}$ for which $(\Pi(f^{(m_l)})r)_i$ converges, it also follows that every limit point of $(\Pi(f^{(m)})r)_i$ equals $(\Pi(f)r)_i$, which guarantees the continuity of $(\Pi(f)r)_i$.

To show that B is a set of reference states for policy f we consider any two states $i, j \in B$. For any state $h \in E$ which is accessible from i under policy f, implying that $P_{ih}^{n_0}(f) > 0$ for some $n_0 > 0$, we have by the pointwise continuity of $P^{n}(f)$ that $P_{ih}^{n_0}(f^{(m_k)}) > 0$, for k large enough, say $k > k_0$. Since B is a set of reference states for $f^{(m_k)}$, $k = 1, 2, \ldots$, it holds that $P_{jh}^{n}(f^{(m_k)}) = 0$, for all $n > 0$, implying that $P_{jh}^{n}(f) = 0$ for all $n > 0$, indicating that h cannot be accessed from j under policy f. This implies that for policy f the class of states containing or accessible from i is different from the class containing or accessible from j. It now follows that $\lim_{k\to\infty} \nu(f^{(m_k)}) = |B| \leqslant \nu(f)$, where $\nu(f)$ denotes the number of classes under policy f. On the other hand, as by Assumption 2 the μ-*GRRS* property is valid for this set B, we have

$$(6.3) \qquad \left({}_{B}P^{n}(f^{(m_k)})\mu\right)_i \leqslant c\beta^{n}\mu_i, \qquad n = 1, 2, \ldots, \quad i \in E, \quad k = 1, 2, \ldots,$$

which again implies that

$$(6.4) \qquad \left({}_{B}P^{n}(f)\mu\right)_i \leqslant c\beta^{n}\mu_i, \qquad n = 1, 2, \ldots, \quad i \in E.$$

Similarly as for Lemma 3.1 one can now prove that $F_{tB}(f) = 1$, which implies that B contains a set of reference states for policy f. Hence $\nu(f) \leqslant |B|$. We conclude that $\nu(f) = |B|$ and consequently that B is a set of reference states for policy f.

To show that, for each state i, $(\Pi(f^{(m_k)})r)_i$ converges and that the limit is equal to $(\Pi(f)r)_i$, we first notice that Lemma 3.3 provides explicit expressions for $(\Pi(f^{(m_k)})r)_i$. Remark further that by using B as set of reference states for policy f in Lemma 3.3 we obtain similar explicit expressions for $(\Pi(f)r)_i$ and that it suffices to show that $\sum_{n=0}^{\infty} {}_B P^n(f^{(m_k)})r$ converges to $\sum_{n=0}^{\infty} {}_B P^n(f)r$. This last fact follows directly from (6.4) and the fact that (6.3) holds uniform for all $f^{(m_k)}$, $k = 1, 2, \ldots$. This completes the proof.

PROOF OF COROLLARY 3.11. The proof follows from Theorem 4.7 together with the concluding remark of §5 of Dekker and Hordijk (1988). Indeed Condition 5 there is Condition A here and in §2 we proved that Assumption 2 implies Condition A. Condition 6(i) there is Assumption 3(i) of this paper. The boundedness of $\|D(f)\|_\mu$ as stated in Condition 6(iii) was shown in Theorem 3.5(iv). The other part of Condition 6(iii) is a consequence of Assumption 3(ii). Finally, the μ-continuity of $D(f)$ is the assertion of Theorem 3.10(ii).

PROOF OF LEMMA 4.1. Since all states communicate and are recurrent under all policies, we can take state 0 as reference state for all policies.

We will establish μ-geometric recurrence by showing that $\sum_{j \neq 0} P_{ij}(f)\mu_j \leqslant \beta\mu_i$ for all $i \in E$, $f \in F$ and some $\beta < 1$. For state 0 the action can be represented by the parameter q and we have

$$\frac{1}{\mu_0} \sum_{j \neq 0} P_{0j}(q)\mu_j = e^{(x-1)\lambda q} - e^{-\lambda q}. \tag{6.5}$$

For $x = 1$ the right-hand side equals $1 - e^{-\lambda q} < 1$ for all $a \leqslant q \leqslant b$, hence it will be clear that for x close enough to 1 there is a $\beta < 1$ such that $e^{(x-1)\lambda q} - e^{-\lambda q} \leqslant \beta$ for all $a \leqslant q \leqslant b$. For $i \neq 0$ we have

$$\frac{1}{\mu_i} \sum_{j=i-1}^{\infty} P_{ij}(q)\mu_j = \frac{1}{x} e^{(x-1)\lambda q} + \left(1 - \frac{1}{x}\right) e^{(x-1)\lambda q - \theta}.$$

Let us denote the right-hand side by $f(x, q)$. It is clear that $\sup_{a \leqslant q \leqslant b} f(x, q) = f(x, b)$ for all $x > 1$. Let

$$g(x) \equiv f(x, b) = \frac{1}{x} e^{(x-1)\lambda b} + \left(1 - \frac{1}{x}\right) e^{(x-1)\lambda b - \theta}, \qquad x > 1,$$

$$g(1) = 1, \quad g'(1) = -1 + \lambda b + e^{-\theta}.$$

Since we assumed that $\lambda b + e^{-\theta} < 1$, we have $g'(1) < 0$ and for small $x > 1$ therefore $g(x) < 1$. Finally we choose x and β so that the requirement for $i = 0$ is also fulfilled.

PROOF OF LEMMA 5.2. It will be obvious that condition μ-*GRRS*(M) implies both conditions μ-*URRS*(M) and μ-*BRRS*(M), so it suffices to show that condition μ-*URRS*(M) implies condition μ-*GRRS*(M) and finally that condition μ-*BRRS*(M) implies condition μ-*URRS*(M).

Suppose therefore that condition μ-*URRS*(M) holds for some constants $n_0 > 0$, $c_2 > 0$ and $\beta < 1$. For $kn_0 \leqslant m < (k+1)n_0$, $k \geqslant 0$ we have

$$\begin{aligned}\left\|{}_{B(f)}P^m(f)\right\|_\mu &\leqslant \left\|{}_{B(f)}P^{kn_0}(f)\right\|_\mu \left\|{}_{B(f)}P^{m-kn_0}(f)\right\|_\mu \\ &\leqslant \beta^k \sup_{1\leqslant l\leqslant n_0} \left\|P^l(f)\right\|_\mu.\end{aligned}$$

Let $\theta = \beta^{1/n_0}$, then $\theta < 1$ and $\beta^k \leqslant \theta^m \beta^{-1}$. Furthermore, $\sup_{1\leqslant l\leqslant n_0} \|P^l(f)\|_\mu \leqslant (c_2)^{n_0}$. Hence

$$\left\|{}_{B(f)}P^m(f)\right\|_\mu \leqslant \left[(c_2)^{n_0}\beta^{-1}\right]\theta^m,$$

which establishes condition μ-*GRRS*(M), with constants $c_1 = (c_2)^{n_0}\beta^{-1} > 0$, and $\theta < 1$.

Suppose condition μ-*BRRS*(M) holds for some constant $c_3 > 0$. Let $y(f) = \sum_{m=0}^{\infty} {}_{B(f)}P^m(f)\mu$. By condition μ-*BRRS*(M) we have $\mu \leqslant y(f) \leqslant c_3\mu$ (in the vector ordering). Since $\mu + {}_{B(f)}P(f)y(f) = y(f)$ we have

$${}_{B(f)}P(f)y(f) = y(f) - \mu \leqslant \left(1 - \frac{1}{c_3}\right)y(f).$$

Choose a $\beta < 1$ and let n_0 be such that $(1 - 1/c_3)^{n_0}c_3 < \beta$, then

$${}_{B(f)}P^{n_0}(f)\mu \leqslant {}_{B(f)}P^{n_0}(f)y(f) \leqslant \left(1 - \frac{1}{c_3}\right)^{n_0} y(f) \leqslant \beta\mu.$$

As n_0 is independent of f, the first part of condition μ-*URRS*(M) has been established for this n_0 and β. Since

$$\begin{aligned}\frac{1}{\mu_i}\sum_{j\in E} P_{ij}(f)\mu_j &\leqslant \frac{1}{\mu_i}\left[\sum_{j\in E} {}_{B(f)}P_{ij}(f)\mu_j + \sum_{j\in B(f)} \mu_j\right] \\ &\leqslant c_3 + \sup_{i\in E}\frac{1}{\mu_i}\sum_{j\in M}\mu_j < \infty, \qquad f \in F,\end{aligned}$$

the second part of condition μ-*URRS*(M) has been established as well for

$$c_2 = c_3 + \sup_{i\in E}\frac{1}{\mu_i}\sum_{j\in M}\mu_j.$$

PROOF OF LEMMA 5.3. Consider next the positive dynamic programming problem with $\tilde{r}_i(a)$ equal to μ_i and

$$\tilde{P}_{ij}(a) = \begin{cases} P_{ij}(a), & j \in M, \\ 0, & j \notin M. \end{cases}$$

It is well known (see Hordijk 1974) that the value vector $\tilde{\mu}$ of this problem satisfies

$$\mu_i + \sum_{j \notin M} P_{ij}(a)\tilde{\mu}_j \leqslant \tilde{\mu}_i, \qquad i \in E, \quad a \in A(i) \quad \text{and}$$

$$\tilde{\mu} = \sup_{f \in F} \sum_{k=0}^{\infty} {}_M P^k(f)\mu.$$

Relation (5.1) implies that $\tilde{\mu} \leqslant \tilde{c}\mu$ for a finite $\tilde{c}$. Hence

$$\sum_{j \in M} P_{ij}(a)\tilde{\mu}_j \leqslant \left(1 - \frac{1}{\tilde{c}}\right)\tilde{\mu}_i.$$

Note that $\tilde{\mu}_i \geqslant \mu_i \geqslant 1$, thus $\tilde{\mu}$ is a bounding vector. The new contraction modulus $\tilde{\beta}$ is equal to $(1 - 1/\tilde{c})$.

Acknowledgements. The authors are grateful to the referee for his comments.

References

Chung, K. L. (1960). *Markov Chains with Stationary Transition Probabilities*. Springer Verlag, Berlin.

Dekker, R. (1985). Denumerable Markov Decision Chains: Optimal Policies for Small Interest Rates. Ph.D. thesis, Univ. of Leiden.

_____ and Hordijk, A. (1988). Average, Sensitive and Blackwell Optimal Policies in Denumerable Markov Decision Chains with Unbounded Rewards. *Math. Oper. Res.* **13** 395–421.

_____ and _____ (1991). Denumerable Semi-Markov Decision Chains with Small Interest Rates. *Ann. Oper. Res.* **28** 185–212.

_____, _____ and Spieksma, F. (1990). On the Relation Between Recurrence and Geometric Ergodicity Conditions in Denumerable Markov Decision Chains. Submitted for publication.

Federgruen, A., Hordijk, A. and Tijms, H. C. (1978). A Note on Simultaneous Recurrence Conditions on a Set of Denumerable Stochastic Matrices. *J. Appl. Probab.* **15** 842–847.

_____, _____ and _____. (1979). Denumerable State Semi-Markov Decision Processes with Unbounded Costs, Average Cost Criterion. *Stochast. Process. Appl.* **9** 223–235.

_____, Schweitzer, P. J. and Tijms, H. C. (1983). Denumerable Undiscounted Semi-Markov Decision Processes with Unbounded Rewards. *Math. Oper. Res.* **8** 298–313.

Hordijk, A. (1974). Dynamic Programming and Markov Potential Theory. Math. Centre. Tract. no. 51, Amsterdam.

_____ and Dekker, R. (1983). Denumerable Markov Decision Chains: Sensitive Optimality Criteria. *Oper. Res. Proc.* 1982. Springer-Verlag, Berlin.

_____ and Kallenberg, L. C. M. (1984). Transient Policies in Discrete Dynamic Programming: Linear Programming Including Suboptimality Tests and Additional Constraints. *Math. Programming* **30** 46–70.

_____ and Sladky, K. (1977). Sensitive Optimality Criteria in Countable State Dynamic Programming. *Math. Oper. Res.* **2** 1–14.

_____ and Spieksma, F. (1989). On Ergodicity and Recurrence Properties of a Markov Chain with an Application to an Open Jackson Network. *Adv. in Appl. Probab.* (to appear).

Laserre, J. B. (1988). Conditions for Existence of Average and Blackwell Optimal Stationary Policies in Denumerable Markov Decision Processes. *J. Math. Anal. Appl.* **136** 479–490.

Mann, E. (1985). Optimality Equations and Sensitive Optimality in Bounded Markov Decision Processes. *Optimization* **16** 757–781.

Popov, N. N. (1977). Conditions for Geometric Ergodicity of Countable Markov Chains. *Soviet Math. Dokl.* **18** 676–679.

Schäl, M. (1987). Estimation and Control in Discounted Dynamic Programming. *Stochastics* **20** 51–71.

Schweitzer, P. J. (1982). Solving MDP Functional Equations by Lexicographic Optimization. *R.A.I.R.O.-Operations Research* **16** 91–98.

Spieksma, F. (1991). The Existence of Sensitive Optimal Policies in Two Multi-Dimensional Queueing Models. *Ann. Oper. Res.* **28** 273–296.

Thomas, L. C. (1980). Connectedness Conditions for Denumerable State Markov Decision Processes. In: R. Hartley, L. C. Thomas, D. J. White (Eds.), *Recent Developments in Markov Decision Processes*. Academic Press, New York, 181–204.

Titchmarsh, E. C. (1939). *The Theory of Functions*. Oxford University Press, Oxford.

Van Hee, K. M. and Wessels, J. (1978). Markov Decision Processes and Strongly Excessive Functions. *Stochast. Process. Appl.* **8** 59–76.

Van Nunen, J. A. E. E. and Wessels, J. (1977). Markov Decision Processes with Unbounded Rewards. In: H. C. Tijms, J. Wessels (Eds.), *Markov Decision Theory*. Math. Centre Tract. 93, 1–24.

Veinott, A. F., Jr. (1969). On Discrete Dynamic Programming with Sensitive Discount Optimality Criteria. *Ann. Math. Statist.* **40** 1635–1660.

Zijm, W. H. M. (1985). The Optimality Equations in Multichain Denumerable State Markov Decision Processes with the Average Cost Criterion: The Bounded Cost Case. *Statist. & Decisions* **3** 143–165.

DEKKER: DEPT. MFTS, SHELL INTERNATIONALE PETROLEUM MIJ B.V., THE HAGUE, THE NETHERLANDS

HORDIJK: INSTITUTE OF APPLIED MATHEMATICS AND COMPUTER SCIENCE, UNIVERSITY OF LEIDEN, LEIDEN, THE NETHERLANDS