Annals of Operations Research 28 (1991) 185–212

DENUMERABLE SEMI-MARKOV DECISION CHAINS WITH SMALL INTEREST RATES *

Rommert DEKKER ** and Arie HORDIJK
Department of Mathematics and Computer Science, University of Leiden, P.O. Box 9512, 2300 RA Leiden, The Netherlands

In this paper we investigate denumerable state semi-Markov decision chains with small interest rates. We consider average and Blackwell optimality and allow for multiple closed sets and unbounded immediate rewards. Our analysis uses the existence of a Laurent series expansion for the total discounted rewards and the continuity of its terms. The assumptions are expressed in terms of a weighted supremum norm. Our method is based on an algebraic treatment of Laurent series; it constructs an appropriate linear space with a lexicographic ordering. Using two operators and a positiveness property we establish the existence of bounded solutions to optimality equations. The theory is illustrated with an example of a K-dimensional queueing system. This paper is strongly based on the work of Denardo [11] and Dekker and Hordijk [7].

1. Introduction

The history of semi-Markov decision chains (abbreviated by SMDC) started in the beginning of the sixties, following the introduction of the semi-Markov process. In literature, the term Markov renewal program, which is the decision variant of the Markov renewal process, is also used, the difference lying only in the description of the underlying process. The theory for SMDC was, and is, closely related to that for (discrete-time) Markov decision chains (MDC), since most concepts and ideas were first developed for MDC and then generalized to SMDC; in the same way this paper is a generalization of Dekker and Hordijk [7].

Initially, the theory for SMDC was developed for models with a finite state space and finite action sets. Several optimality criteria were investigated, such as discount and average optimality, cf. De Cani [6], Howard [20], Jewell [21] and Schweitzer [32]. Subsequently, considerable effort was spent on the generalization to models with a denumerable state space and compact action sets. The extension of discount optimality to the denumerable state and compact action case is quite straightforward, cf. Blackwell [3]. For average optimality however, the denumer-

* This research has partially been sponsored by the Netherlands Organization for Scientific Research (NWO).
** Present address: Dept. MFTS, Shell Internationale Petroleum Mij B.V., The Hague, The Netherlands.

able state space gave complications concerning the asymptotic behaviour of the embedded Markov chain. The first results required a simple Markov chain structure such as ergodicity or later unichainedness, both together with bounded rewards, cf. Derman [13], Ross [27] for the discrete-time case and Ross [28], Hordijk [18], Federgruen and Tijms [16] for the semi-Markov case; the last two papers also allowed compact action sets. This was extended to the unbounded rewards case by Hordijk [18], Federgruen, Hordijk and Tijms [14] for the unichain case and by Federgruen, Schweitzer and Tijms [15] for an extension of the communicating case, the communicating case first being set up and investigated for the finite state compact action model by Bather [1]. However, the theory for the denumerable state model remained mathematically unsatisfactory since it did not cover the general finite case. Together with Deppe [12], this paper investigates the denumerable multichain case for the SMDC. For the denumerable multichain MDC case we refer to Zijm [41,42], Mann [24], Dekker and Hordijk [7,8], and Lasserre [23].

In spite of its wide use, average optimality is a rough and insensitive criterion. More sensitive criteria such as n-discount and Blackwell optimality were developed and established. This was done by Blackwell [2,3], Veinott [40], Denardo and Miller [10] and Miller and Veinott [26] for the MDC case and Miller [25] and Denardo [11] for the SMDC case, the latter also establishing a Laurent series expansion for the discounted rewards, where Blackwell optimality is a combination of average optimality and discount optimality for small interest rates. It implies n-discount optimality for all n. In Hordijk and Sladky [19] and Dekker and Hordijk [7] these results were generalized to the denumerable case MDC, whereas this paper contains the extension to the SMDC case.

The outline of this paper is as follows. In section 2 the model is introduced and a Laurent series expansion is derived for the total expected discounted rewards. Section 3 contains an algebraic analysis of these Laurent series and establishes solutions to optimality equations. In section 4 we illustrate our theory with a specific K-dimensional queueing system. We consider a single server queueing network in which there are K service facilities, one server and a random routing of the jobs. In Klimov [22] the structure of the priority rule with minimal average cost is determined. The cost function consists of a linear holding cost. The service time distribution is assumed to be general with finite first and second moment. In Sennott [34] this model is one of the applications and the existence of a deterministic average optimal policy is shown. In section 4 we prove for this model that the Blackwell optimality equations have a unique solution which equals the Laurent expansion of a Blackwell optimal policy. This is a stronger result. However, we must assume that the service distributions have tail probabilities which tend exponentially fast to zero, but this is not a strong assumption as most distribution classes used in practice have this property.

In the last years the existence and structure of optimal policies in queueing systems have been studied in Weber and Stidham [39] and Stidham and Weber

[36]. In Sennott [33] and Sennott [34] existence theorems are applied to queueing systems.

This paper is a revision and extension (section 4) of chapter 2 of Dekker [9]. In this paper we restrict ourselves to stationary deterministic policies. From a recent result of Schäl [31] it follows that our Blackwell optimal policy is at least strong average optimal in the class of all policies.

2. The model and the Laurent series expansion for the total discounted rewards

Consider a dynamic system which is observed at timepoints T_n, $n = 0, 1, \ldots$ with $T_0 = 0$. The time between successive observations is a random variable taking values in $[0, \infty)$. When the system is observed at timepoint T_n, it is characterized by a state, say S_n. We assume the *state space* E to be denumerable and we let $E = \{1, 2, \ldots\}$. In each state i there is a set of *available actions* $A(i)$ for controlling the system. When state i is observed, an action $D(i)$ is chosen from $A(i)$, which remains fixed up to the next observation. The *transition probabilities* are assumed to be time independent. Let

$$Q_{ij}^{a}(x) \equiv P\big(S_{n+1} = j,\ T_{n+1} \leqslant t + x \mid S_n = i,\ T_n = t,\ D(i) = a\big) \tag{2.1}$$

be the joint probability that the next observed state is j and that the transition time is not greater than x, given observation of state i and given decision $D(i) = a$. Let $R_i^a(x)$ denote the expectation of the income during the time interval $[0, \min(x, T_1)]$, given $S_0 = i$ and $D(i) = a$, that is, the expected income until the earlier of time x and the first transition. Let $F \equiv \prod_{i=1}^{\infty} A(i)$ be the set of *decision rules*. We shall restrict ourselves to the class of *stationary policies*, i.e. policies of the type $f^\infty \equiv \{f, f, \ldots\}$. However, in order to keep the notation short, we write f instead of f^∞.

We shall use the following matrix and vector notation. Let $Q^f(x)$ denote the matrix with $Q_{ij}^{f(i)}(x)$ as (i, j)th element and let $R^f(x)$ denote the vector with $R_i^{f(i)}(x)$ as ith element. Further, let the matrix $P(f)$ be defined as $P_{ij}(f) \equiv Q_{ij}^{f(i)}(\infty)$. Note that $P(f)$ is the transition matrix of the embedded Markov chain. Let e denote the vector with components equal to one. $P^k(f)$ is defined by $P^k(f) \equiv P(f)P^{k-1}(f)$, $k \geqslant 1$, where $P^0(f) \equiv I$, the *identity matrix*. We will make the following standard assumption.

ASSUMPTION 1

(a) $\sum_{j=1}^{\infty} P_{ij}(f) = 1$ for all $i \in E$, $f \in F$.

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

(c) $R_i^f(t)$ is of bounded variation as a function of t, for all $i \in E$, $f \in F$.

The bounded variation of $R_i^f(t)$ means that there exist two nondecreasing nonnegative functions denoted by $(R_i^f)^+(t)$ and $(R_i^f)^-(t)$ respectively, such that $(R_i^f)(t) = (R_i^f)^+(t) - (R_i^f)^-(t)$. Note that under assumption 1(b) F is a compact metric set in the product topology. Let $M_{ij}^f(t)$ denote the number of timepoints T_n, $n = 0, 1, 2, \ldots$ such that $S_n = j$ and let $U_i^f(t) = \sum_j M_{ij}^f(t)$. In words, $M_{ij}^f(t)$ and $U_i^f(t)$ denote the expectation of the number of "transitions" to state j and the total number of transitions respectively made in the time interval $[0, t]$ given $S_0 = i$ and policy f.

As reward criterion we use $V_i^f(t)$, being the total expected income during the time interval $[0, t]$ given $S_0 = i$ and policy f. $V_i^f(t)$ may be undefined or infinite in a denumerable state space with unbounded immediate rewards. In order to exclude this, we assume that the immediate rewards are bounded by a vector μ and that the matrix M has certain properties with respect to this vector. This vector is then said to be a bounding vector. This leads to assumptions in terms of this bounding vector and to an operator-theoretical approach based on the linear space induced by the bounding vector. This approach is, however, somewhat stronger than working with Liapunov functions, cf. Hordijk [18]. It has been shown that the existence of a first order Liapunov function is almost sufficient for average optimality; however, for Blackwell optimality, Liapunov functions of any order have to be assumed and this is almost equivalent to the existence of a bounding vector satisfying our assumptions.

For any vector μ on E with $\mu_i \geqslant 1$, for all $i \in E$, we can define a *weighted supremum norm* $\| \cdot \|_\mu$ as follows

$$\| r \|_\mu \equiv \sup_{i \in E} \frac{| r_i |}{\mu_i},$$

for any vector r on E. Its corresponding matrix or operator norm is defined as follows:

$$\| A \|_\mu := \sup_{r : \| r \|_\mu \neq 0} \frac{\| Ar \|_\mu}{\| r \|_\mu} = \sup_{i \in E} \frac{1}{\mu_i} \sum_j | A_{ij} | \mu_j,$$

where A: $E \times E \to \mathbb{R}$ is any matrix function. The following assumption states precisely the properties we require.

ASSUMPTION 2

There exists a vector μ on E with $\mu_i \geqslant 1$ for all $i \in E$ and constants c_0^+, c_0^-, c_1, c_2 such that for all $f \in F$:
(a) $\sup_{x \geqslant 0} \| (R^f)^+(x) \|_\mu \leqslant c_0^+$, $\sup_{x \geqslant 0} \| (R^f)^-(x) \|_\mu \leqslant c_0^-$.
(b) $\| M^f(t) \|_\mu \leqslant c_1 + c_2 t$, for all $t > 0$.

We call the vector μ a *bounding vector*. A vector r, matrix A on E are called μ-bounded if $\| r \|_\mu < \infty$, $\| A \|_\mu < \infty$ respectively. We denote the set of μ-bounded vectors, matrices on E by V^μ, M^μ respectively. Assumption 2(a) is

stated in this way to guarantee bounded variation of the quantities to be considered later. It implies $\sup_{x \geqslant 0} \| R^f(x) \|_\mu \leqslant c_0$, where $c_0 \equiv c_0^+ + c_0^-$.

In the following, we shall use the Lebesgue–Stieltjes integral $\int_a^b g(x)\, \mathrm{d}F(x)$, where $g(x)$ is a Borel measurable function and $F(x)$ is of bounded variation, cf. Royden [29, p. 263].

The *convolution* of $Q^f(t)$ with itself or with $R^f(t)$, denoted by $(Q^f)^{*2}(t)$ and $(Q^f * R^f)(t)$ respectively, is defined by

$$(Q^f)_{ij}^{*2}(t) \equiv \int_0^t \sum_k Q_{kj}^f(t-x)\, \mathrm{d}Q_{ik}^f(x), \qquad t \geqslant 0,\ i,\ j \in E,\ f \in F \tag{2.2}$$

and

$$(Q^f * R^f)_i(t) \equiv \int_0^t \sum_k R_k^f(t-x)\, \mathrm{d}Q_{ik}^f(x), \quad t \geqslant 0,\ i,\ j \in E,\ f \in F. \tag{2.3}$$

Using standard arguments it can be seen that $(Q^f * R^f)_i(t)$ is again of bounded variation. Note that $(Q^f)_{ij}^{*2}(t)$ is the probability that the second transition is made to state j before time t, given $S_0 = i$ and policy f. Analogously, $(Q^f * R^f)_i(t)$ is the expectation of the reward between the first transition and the earlier of the second transition and time t, given $S_0 = i$ and policy f. Denote the n-fold convolution of $Q^f(t)$ with itself by $(Q^f)^{*n}(t)$. Observe further that

$$M_{ij}^f(t) = \sum_{k=0}^{\infty} (Q^f)_{ij}^{*k}(t), \quad t \geqslant 0,\ i,\ j \in E,\ f \in F, \tag{2.4}$$

where $(Q^f)_{ij}^{*0}(t)$ is equal to 1, 0 if $i = j$ respectively $i \neq j$. A formal definition of $V_i^f(t)$ can now be given by

$$V_i^f(t) = \sum_{k=0}^{\infty} \left((Q^f)^{*k} * R^f \right)_i(t), \quad t \geqslant 0,\ i \in E,\ f \in F, \tag{2.5}$$

where $((Q^f)^{*0} * R^f)_i(t) = R_i^f(t)$. By (2.4) we have $V_i^f(t) = (M^f * R^f)_i(t)$. It is easy to show that $V_i^f(t)$ is of bounded variation for all $i \in E$ and that its μ-norm is bounded by

$$\| V^f(t) \|_\mu \leqslant c_0(c_1 + c_2 t). \tag{2.6}$$

We shall discount the total rewards with a rate $s > 0$, i.e. we take the Laplace–Stieltjes transform of $V^f(t)$ defined by

$$v_i^f(s) \equiv \int_0^\infty \mathrm{e}^{-st}\, \mathrm{d}V_i^f(t), \quad s > 0,\ i \in E,\ f \in F. \tag{2.7}$$

Its existence follows for all $s > 0$ from (2.6) and the bounded variation of $V^f(t)$. It is easily seen that

$$\| v^f(s) \|_\mu \leqslant c_0\left(c_1 + c_2 s^{-1}\right), \quad s > 0,\ f \in F. \tag{2.8}$$

We also take the Laplace–Stieltjes transforms of $Q_{ij}^f(t)$ and $R_i^f(t)$ and denote them by $q_{ij}^f(s)$ and $r_i^f(s)$ respectively. From (2.5) it easily follows that

$$V_i^f(t) = R_i^f(t) + \sum_j \int_0^t V_j^f(t-x)\, \mathrm{d}Q_{ij}^f(x), \quad t \geqslant 0,\ i \in E,\ f \in F. \tag{2.9}$$

Taking Laplace–Stieltjes transforms of (2.9) is possible under our assumptions and yields (in matrix notation)

$$v^f(s) = r^f(s) + q^f(s)v^f(s), \quad s > 0,\ f \in F. \tag{2.10}$$

This equation is the key to our analysis. Our first objective is to give conditions implying that $v^f(s)$ has a Laurent series expansion. Since $V_i^f(t) = O(t)$ as $t \to \infty$ for all $i \in E$ (i.e. $\limsup_{t\to\infty} t^{-1} |V_i^f(t)| < \infty$), we expect to find $v_i^f(s) = O(s^{-1})$ as $s \to 0$ and hence a Laurent series expansion for $v_i^f(s)$ which starts with a s^{-1} term. First of all, we shall prove that (2.10) has unique μ-bounded solutions under assumptions 1 and 2. Secondly we shall show that there are μ-bounded vectors $W^{(k)}(f)$, $k = -1, 0, 1, \ldots$ such that $v^f(s) = \sum_{k=-1}^{\infty} W^{(k)}(f)s^k$ is the μ-bounded solution to (2.10). Let

$$m^f(s) = \sum_{n=0}^{\infty} \left[q^f(s)\right]^n. \tag{2.11}$$

From assumption 2(b), eq. (2.4) and the nondecreasingness of $M_{ij}^f(t)$ it follows that the Laplace–Stieltjes transform of $M_{ij}^f(t)$ exists and is equal to $m_{ij}^f(s)$. Taking μ-norms, we find $\| m^f(s) \|_\mu \leqslant c_1 + c_2 s^{-1} < \infty$ for all $s > 0$.

LEMMA 2.1

Under assumptions 1 and 2, $v^f(s)$ is the unique μ-bounded solution to the set of equations

$$v = r^f(s) + q^f(s)v, \quad s > 0,\ f \in F,\ v \in V^\mu. \tag{2.12}$$

Proof

By (2.10) we only have to show the uniqueness of μ-bounded solutions of (2.12). Suppose v and w are μ-bounded solutions to (2.12), i.e. both $v = r^f(s) + q^f(s)v$ and $w = r^f(s) + q^f(s)w$. Subtracting these equations yields $v - w = [q^f(s)][v - w]$ and iterating this equality n times gives $v - w = [q^f(s)]^n[v - w]$, for $n = 1, 2, \ldots$. Adding these equations and taking μ-norms yields for every $m \in \mathbb{N}$ that $\| v - w \|_\mu \leqslant m^{-1} \left\| \sum_{n=0}^{m} \left[q^f(s)\right]^n \right\|_\mu \| v - w \|_\mu$. Since $m^f(s) = \sum_{n=0}^{\infty} [q^f(s)]^n$ and $\| m^f(s) \|_\mu < \infty$ for all $s > 0$ and $f \in F$ we find that $\| w - w \|_\mu = 0$. □

Note that we can write

$$v^f(s) = m^f(s)r^f(s), \quad s > 0,\ f \in F. \tag{2.13}$$

Since $q_{ij}^f(s)$ and $r_i^f(s)$ are Laplace–Stieltjes transforms, they are analytic in the open halfplane $\mathrm{Re}(s) > 0$ (when considered as a function of complex s); however, this is not sufficient for our analysis. Therefore we make the following assumption.

ASSUMPTION 3

There exist positive constants A, B and c_3, such that for all $f \in F$

(a) $$Q_{ij}^{(k)}(f) \equiv \int_0^\infty \frac{x^k}{k!}\, \mathrm{d}Q_{ij}^f(x)$$

exists for all $i,\ j \in E$ and $k = 0, 1, \ldots$ and $\| Q^{(k)}(f) \|_\mu \leqslant A(c_3)^k,\ k = 0, 1, \ldots;$

(b) $$R_i^{(k)}(f) \equiv \int_0^\infty \frac{x^k}{k!}\, \mathrm{d}R_i^f(x)$$

exists for all $i \in E$ and $k = 0, 1, \ldots$ and $\| R^{(k)}(f) \|_\mu \leqslant B(c_3)^k,\ k = 0, 1, \ldots.$

Assumption 3 states in fact that all moments of the times to transition and rewards have to exists, which is the case for most distribution classes used in practice. More precisely, under assumption 3 the following expansions exist for $q^f(s)$ and $r^f(s)$ respectively:

$$q^f(s) = \sum_{k=0}^{\infty} Q^{(k)}(f)(-s)^k, \quad f \in F,\ |s| \leqslant (c_3)^{-1}, \tag{2.14}$$

$$r^f(s) = \sum_{k=0}^{\infty} R^{(k)}(f)(-s)^k, \quad f \in F,\ |s| \leqslant (c_3)^{-1}. \tag{2.15}$$

Hence, $q_{ij}^f(s)$ and $r_i^f(s)$ are analytic for all $i,\ j \in E,\ f \in F$ on the set $\{ s \in \mathbb{C} \mid 0 \leqslant |s| < (c_3)^{-1} \}$. Assumption 3 is most natural for establishing a Laurent series expansion for $v_i^f(s)$ around zero, starting with a term s^{-1}. It implies that $v_i^f(s)$ is analytic on some disc around zero except zero itself.

It will be clear that the properties of the embedded Markov chain are important for the existence of average and/or Blackwell optimal policies. We shall therefore make a detailed analysis of the embedded chain, in which we follow Dekker and Hordijk [7].

It is well known, cf. Chung [4], that the Cesaro limit of $P^k(f)$ always exists, i.e.

$$\Pi_{ij}(f) = \lim_{N \to \infty} \frac{1}{N+1} \sum_{k=0}^{N} P_{ij}^k(f), \tag{2.16}$$

exists for all $i,\ j,\ \in E,\ f \in F$. Let $\Pi(f)$ denote the matrix with $\Pi_{ij}(f)$ as $(i,\ j)$th element. $\Pi(f)$ is also called the *stationary matrix* since the following equalities hold

$$\Pi(f) = \Pi(f)P(f) = P(f)\Pi(f) = \Pi(f)\Pi(f), \quad f \in F. \tag{2.17}$$

Since $\Pi(f)$ contains the information on the long-term behaviour of the Markov chain, it is basic for the average rewards over an infinite horizon. In addition, we are also interested in sensitive optimality criteria, dealing with deviations from the average rewards. Therefore, following Dekker and Hordijk [7], we assume

ASSUMPTION 4

(a) $D_{ij}(f) \equiv \lim_{\alpha \uparrow 1} \sum_{k=0}^{\infty} \alpha^k [P_{ij}^k(f) - \Pi_{ij}(f)]$ exists for all $i, j \in E, f \in F$.

(b) For every $f \in F$ the following relations hold for the matrix $D(f)$:

$$\Pi(f) D(f) = D(f) \Pi(f) = 0, \tag{2.18}$$

$$[I - P(f)] D(f) = D(f)[I - P(f)] = I - \Pi(f), \tag{2.19}$$

where 0, I denote the zero and identity matrix respectively.

(c) There exists a constant c_4 such that $\| D(f) \|_\mu \leqslant c_4$, for every $f \in F$.

The matrix $D(f)$ is called the *deviation matrix*. To facilitate verification of assumption (4) sufficient recurrence conditions are provided in Dekker and Hordijk [8]. From assumptions 1 and 4 it can be shown (see remark 1 of Dekker and Hordijk [7]) that

$$\lim_{N \to \infty} \frac{1}{N+1} \sum_{k=0}^{N} P^k(f) r = \Pi(f) r, \tag{2.20}$$

for all μ-bounded vectors r and all $f \in F$. Consequently, $\sum_j \Pi_{ij}(f) = 1$ for all $i \in E$ and $f \in F$.

Before we proceed with our analysis, we first recapitulate some facts of Markov chains. A nonempty set C is called *closed under P(f)* if $P_{ij}(f) = 0$ for all $i \in C$, $j \notin C$. It is called *minimal closed under P(f)* if it does not contain a proper and closed subset. The foregoing implies that in the embedded Markov chains all minimal closed sets are positive recurrent and all inessential states are transient. Let $\nu(f)$ denote the number of minimal closed sets of the Markov chain under $P(f)$ and denote these sets by $C_1, \ldots, C_{\nu(f)}$. Note that $\nu(f)$ can be infinite. Let $T(f)$ denote the set of transient states and $R(f)$ the set of positive recurrent states under policy f, then $R(f) = C_1 \cup \ldots \cup C_{\nu(f)}$ and $E = R(f) \cup T(f)$. We denote by $F_{iC_n}(f)$ the probability that the set C_n is reached under policy f when the system starts in state i.

Assumption 3(a) states for $k = 0$ that $Q^{(0)}(f) = P(f)$ and that $\| P(f) \|_\mu \leqslant A$ for all $f \in F$. Combining this with (2.19) and assumption 4(c) we see that $\Pi(f)$ has a bounded (in the policy space) μ-norm, i.e.

$$\| \Pi(f) \|_\mu \leqslant 1 + c_4(1 + A), \quad f \in F. \tag{2.21}$$

Let $\tau_i(f)$ denote the expected time until transition in state i under policy f. Note that $\tau_i(f) = \sum_j Q_{ij}^{(1)}(f)$ and that the vector $\tau(f)$ is also μ-bounded. Using (2.21) we see that $\Pi(f)\tau(f)$ also has a bounded μ-norm.

Let $\tau_{jj}(f)$ denote the mean recurrence time in state j under policy f. Consider a set $C \subset E$ which is minimal closed under $P(f)$. According to Cinlar [5, theorem 10.4.3] we have, for any state $i \in C$,

$$\tau_{ii}(f) = \frac{\sum_{j \in C} \Pi_{ij}(f)\tau_j(f)}{\Pi_{ii}(f)}. \tag{2.22}$$

From the preceding it follows that C is positive recurrent in the embedded Markov chain, so $\Pi_{ii}(f) > 0$. Since in addition $\sum_{j \in C} \Pi_{ij}(f)\tau_j(f) < \infty$, we find that $\tau_{ii}(f) < \infty$ and that C is also a class of positive recurrent states in the semi-Markov process. Taking inverses in (2.22) and summing over all states in C yields

$$\sum_{i \in C} \frac{1}{\tau_{ii}(f)} = \frac{1}{\sum_{j \in C} \Pi_{jj}(f)\tau_j(f)}. \tag{2.23}$$

In our analysis we shall need the boundedness of the right-hand side of (2.23) both in f and in C, as stated in the next lemma.

LEMMA 2.2

Under assumptions $1, \ldots, 4$ we have for every $f \in F$ and all minimal closed sets C under $P(f)$

$$\frac{1}{\sum_{i \in C} \Pi_{ii}(f)\tau_i(f)} \leqslant c_2. \tag{2.24}$$

Proof

Consider a policy f and a set C minimal closed under $P(f)$. It follows from elementary renewal theory that, for all $i, j \in C$,

$$\lim_{t \to \infty} \frac{1}{t} M_{ij}^f(t) = \frac{1}{\tau_{jj}(f)}.$$

Accordingly,

$$\lim_{t \to \infty} \frac{1}{t} \sum_{j \in C} M_{ij}^f(t) = \sum_{j \in C} \frac{1}{\tau_{jj}(f)} = \frac{1}{\sum_{j \in C} \Pi_{jj}(f)\tau_j(f)}.$$

On the other hand, assumption 2(b) states that, for all $i \in C$,

$$\sum_{j \in C} M_{ij}^f(t) \frac{\mu_j}{\mu_i} \leqslant c_1 + c_2 t.$$

Let $\mu_0 \equiv \inf_{i \in C} \mu_i$, then obviously $\mu_0 \geqslant 1$. Hence for every $\epsilon > 0$ there is an $i_\epsilon \in E$ such that $\mu_{i_\epsilon} \leqslant \mu_0 + \epsilon$, implying for all $i \in C$,

$$\frac{\mu_i}{\mu_{i_\epsilon}} \geqslant \frac{\mu_0}{\mu_0 + \epsilon}.$$

This yields

$$\frac{1}{\sum_{j\in C}\Pi_{jj}(f)\tau_j(f)} = \lim_{t\to\infty}\frac{1}{t}\sum_{j\in C}M_{ij}^{f}(t)$$

$$\leqslant \lim_{t\to\infty}\left[\frac{1}{t}\sum_{j\in C}M_{ij}^{f}(t)\frac{\mu_j}{\mu_{i_\epsilon}}\right]\left(\frac{\mu_0}{\mu_0+\epsilon}\right)\leqslant c_2\frac{\mu_0+\epsilon}{\mu_0}.$$

Since this result holds for every $\epsilon > 0$, the lemma follows directly. □

This result is used in the following lemma which characterizes solutions to equations of the type $[I - P(f)]x = b$.

LEMMA 2.3

Let b, c be μ-bounded vectors on E with $\Pi(f)b = 0$ for some $f \in F$. Then

(i) x is a μ-bounded solution to $[I - P(f)]x = b$ if and only if x is a μ-bounded solution to $[I - \Pi(f)]x = D(f)b$;

(ii) the set of equations

$$[I - P(f)]x = b \quad \text{and} \quad \Pi(f)Q^{(1)}(f)x = \Pi(f)c \tag{2.25}$$

has a unique μ-bounded solution x which is given by

$$x = D(f)b + \Pi(f)\left[\frac{\Pi(f)c - \Pi(f)Q^{(1)}(f)D(f)b}{\Pi(f)\tau(f)}\right], \tag{2.26}$$

with the vector quotient taken component by component, i.e.

$$\left[\frac{x}{y}\right]_i \equiv \left[\frac{x_i}{y_i}\right].$$

Proof

(i) Suppose x is a μ-bounded solution to $[I - P(f)]x = b$. Multiplication of this equation by $D(f)$ and the use of (2.19) yields $[I - \Pi(f)]x = D(f)b$. On the other hand, suppose $[I - \Pi(f)]x = D(f)b$, hence $x = D(f)b + \Pi(f)x$ and $[I - P(f)]x = [I - P(f)][D(f) + \Pi(f)x] = [I - \Pi(f)]b = b$.

(ii) Let y be the right-hand side of (2.26). Our assumptions and lemma 2.2 imply that y is μ-bounded. It is easily seen that $[I - P(f)]y = b$. Let

$$z \equiv \Pi(f)\left[\frac{\Pi(f)c - \Pi(f)Q^{(1)}(f)D(f)b}{\Pi(f)\tau(f)}\right].$$

Note that z is a constant on any minimal closed set under $P(f)$, say

$$z_i = d_n = \frac{\left[\Pi(f)c - \Pi(f)Q^{(1)}(f)D(f)b\right]_i}{\left[\Pi(f)\tau(f)\right]_i},$$

for $i \in C_n$, $n = 1, 2, \ldots, \nu(f)$. The values of z in the recurrent states determine the values of z in the transient states, as

$$z_i = \sum_j \Pi_{ij}(f) z_j = \sum_{n=1}^{\nu(f)} F_{iC_n}(f) d_n,$$

for each $i \in T(f)$. Substituting y for x in $\Pi(f)Q^{(1)}(f)x = \Pi(f)c$ yields, after rearranging terms, $\Pi(f)Q^{(1)}(f)z = \Pi(f)c - \Pi(f)Q^{(1)}(f)D(f)b$ and this equality has to be verified only on the recurrent states. Consider some minimal closed set C_n under $P(f)$ and note that it is also minimal closed under $\Pi(f)$ or $Q^{(1)}(f)$. For $i \in C_n$ we have (let e denote the vector with all components equal to one):

$$\begin{aligned}\left[\Pi(f)Q^{(1)}(f)z\right]_i &= \left[\Pi(f)Q^{(1)}(f)d_n e\right]_i = d_n\left[\Pi(f)Q^{(1)}(f)e\right]_i \\ &= d_n\left[\Pi(f)\tau(f)\right]_i = \left[\Pi(f)c\right]_i - \left[\Pi(f)Q^{(1)}(f)D(f)b\right]_i,\end{aligned}$$

which was to be shown.

Suppose w is also a μ-bounded solution to (2.25). Subtraction of the two solutions yields $[I - P(f)][w - y] = 0$ and $\Pi(f)Q^{(1)}(f)(w - y) = 0$. Using the first equality we obtain $(w - y) = \Pi(f)(w - y)$, which implies that $w - y$ is constant on the closed sets under $P(f)$ and that its value in the transient states under $P(f)$ is completely determined by its value on the closed sets. From $\Pi(f)Q^{(1)}(f)(w - y) = 0$ now follows that $w - y = 0$, which provides the proof. □

We are now able to state the main theorem of this section.

THEOREM 2.4

Under assumptions $1, \ldots, 4$ the following holds for every $f \in F$:

$$v^f(s) = \sum_{k=-1}^{\infty} V^{(k)}(f) s^k, \quad 0 < s < d, \tag{2.27}$$

for some constant d independently of f, where $V^{(k)}(f)$, $k = -1, 0, \ldots$ is the unique μ-bounded solution to the set of equations

$$\begin{aligned}&[I - P(f)]V^{(k)}(f) = b^{(k)}(f), \\ &\Pi(f)Q^{(1)}(f)V^{(k)}(f) = \Pi(f)c^{(k)}(f),\end{aligned} \tag{2.28}$$

with $c^{(-1)}(f) \equiv R^{(0)}(f)$, $b^{(-1)}(f) \equiv 0$ and for $k = 0, 1, \ldots$

$$b^{(k)}(f) = c^{(k-1)}(f) - Q^{(1)}(f)V^{(k-1)}(f), \tag{2.29}$$

$$c^{(k)}(f) = (-1)^{(k+1)}R^{(k+1)}(f) + \sum_{j=2}^{k+2} (-1)^j Q^{(j)}(f)V^{(k+1-j)}(f). \tag{2.30}$$

Proof

First observe that $c^{(k)}(f)$ and $b^{(k)}(f)$ depend only on variables $V^{(-1)}(f)$, $V^{(0)}(f), \ldots, V^{(k-1)}(f)$ and moments $Q^{(0)}(f)$ through $Q^{(k+2)}(f)$ and $R^{(0)}(f)$ through $R^{(k+1)}(f)$. Applying this recursively shows that $V^{(k)}(f)$ is defined in terms of moments $Q^{(0)}(f)$ through $Q^{(k+2)}(f)$ and $R^{(0)}(f)$ through $R^{(k+1)}(f)$. By assumption 3, all moments of $Q^f(x)$ and $R^f(x)$ exist and are μ-bounded, hence $c^{(k)}(f)$ and $b^{(k)}(f)$, $k = -1, 0, 1, \ldots$ are also μ-bounded. It then follows from lemma 2.3 that (2.28) has unique μ-bounded solutions $V^{(k)}(f)$ for $k = -1, 0, 1, \ldots$.

Let us write $w_i^f(s) \equiv \sum_{k=-1}^{\infty} V_i^{(k)}(f) s^k$, $s > 0$. Suppose now that $w_i^f(s)$ converges absolutely and uniformly in f, for $0 < s < d$, for some $d > 0$, then the proof proceeds as follows. Substituting $w^f(s)$ and the power series expansion for $q^f(s)$ and $r^f(s)$ respectively in $r^f(s) + [q^f(s) - I]w^f(s)$ yields

$$\sum_{k=0}^{\infty} (-1)^k R^{(k)}(f) s^k + \left[\sum_{n=0}^{\infty} (-1)^n Q^{(n)}(f) s^n\right]\left[\sum_{m=-1}^{\infty} V^{(m)}(f) s^m\right]$$

$$- \sum_{m=-1}^{\infty} V^{(m)}(f) s^m$$

$$= \sum_{k=0}^{\infty} (-1)^k R^{(k)}(f) s^k + \sum_{k=-1}^{\infty} s^k \left[\sum_{n=0}^{k+1} (-1)^n Q^{(n)}(f) V^{(k-n)}(f)\right]$$

$$- \sum_{m=-1}^{\infty} V^{(m)}(f) s^m.$$

Note that all Laurent or power series converge absolutely for s small enough, uniformly for all states. Collecting terms with equal power of s yields

$$\left[-V^{(-1)}(f) + Q^{(0)}(f) V^{(-1)}(f)\right] s^{-1}$$

$$+ \sum_{k=0}^{\infty} s^k \left[(-1)^k R^{(k)}(f) + \sum_{n=0}^{k+1} (-1)^n Q^{(n)}(f) V^{(k-n)}(f) - V^{(k)}(f)\right].$$

The coefficient of the term with s^{-1} is zero by (2.28) with $k = -1$. With respect to the coefficient of the term with s^k we have

$$(-1)^k R^{(k)}(f) + \sum_{n=0}^{k+1} (-1)^n Q^{(n)}(f) V^{(k-n)}(f) - V^{(k)}(f)$$

$$= -\left[I - Q^{(0)}(f)\right] V^{(k)}(f) + (-1)^k R^{(k)}(f)$$

$$+ \sum_{n=2}^{k+1} (-1)^n Q^{(n)}(f) V^{(k-n)}(f) - Q^{(1)}(f) V^{(k-1)}(f)$$

$$= -\left[I - Q^{(0)}(f)\right] V^{(k)}(f) + c^{(k-1)}(f) - Q^{(1)}(f) V^{(k-1)}(f),$$

which is also zero by (2.28) for $k = 0, 1, \ldots$. Hence $w^f(s)$ is a μ-bounded solution to (2.10) and since (2.10) has a unique μ-bounded solution, $w^f(s)$ is equal to $v^f(s)$.

We shall now prove that $w_i^f(s)$ converges absolutely, uniformly in i and f, for s small enough. For this, it is sufficient to show the existence of constants a and b such that

$$\| V^{(n)}(f) \|_\mu \leqslant ab^n, \quad f \in F,\ n = -1, 0, 1, \ldots. \tag{2.31}$$

A detailed proof of this result is rather long, due to the complexity of the expressions for $V^{(n)}(f)$. Therefore we only give an outline of the proof and leave the verification to the reader.

Let $b^{(n)}(f)$ and $c^{(n)}(f)$ be given by (2.29) and (2.30) respectively. We show the existence of constants a and b, not depending on n, such that $\| V^{(n)}(f) \|_\mu \leqslant ab^n$, $\| b^{(n)}(f) \|_\mu \leqslant ab^n$ and $\| c^{(n)}(f) \|_\mu \leqslant ab^n$, $f \in F$, are consequences of $\| V^{(k)}(f) \|_\mu \leqslant ab^k$, $\| b^{(k)}(f) \|_\mu \leqslant ab^k$ and $\| c^{(k)}(f) \|_\mu \leqslant ab^k$, $k = -1, \ldots, n-1$, $f \in F$. Induction to n then gives the desired results. For $n = -1$ it is easy to see that $b^{(-1)}(f) = 0$ and that $\| c^{(-1)}(f) \|_\mu = \| R^{(0)}(f) \|_\mu \leqslant B$ by assumption 3. We further have

$$V^{(-1)}(f) = \Pi(f)\left[\frac{\Pi(f)R^{(0)}(f)}{\Pi(f)\tau(f)}\right],$$

which is μ-bounded by assumptions 2, 3, 4, (2.21) and lemma 2.2. For $n \geqslant 0$ it easily follows from using the same results and the induction hypothesis that $\| b^{(n)}(f) \|_\mu \leqslant ad_1 b^{n-1}$, for some constant d_1. W.r.t. $c^{(n)}(f)$ we have in a similar way

$$\| c^{(n)}(f) \|_\mu \leqslant B(c_3)^{n+1} + aA\left[\sum_{j=0}^{n} (c_3)^2 \left(\frac{c_3}{b}\right)^j\right] b^{n-1} \leqslant ad_2 b^{n-1},$$

for some constant d_2 independent of n, provided that $b > c_3$. Finally, w.r.t. $V^{(n)}(f)$ we have,

$$V^{(n)}(f) = D(f)b^{(n)}(f) + \Pi(f)\left[\frac{\Pi(f)c^{(n)}(f) - \Pi(f)Q^{(1)}(f)D(f)b^{(n)}(f)}{\Pi(f)\tau(f)}\right].$$

Assumptions 2, 3, 4, lemma 2.2 and (2.21) and the results for $b^{(n)}(f)$ and $c^{(n)}(f)$ now imply that $\| V^{(n)}(f) \|_\mu \leqslant ad_3 b^{n-1}$, for some constant d_3, independent of n. Hence, any b larger than $\max(c_3, d_1, d_2, d_3)$ together with a constant a larger than B, such that (2.31) also holds for $n = -1$, fulfills (2.31) for all n. This completes the proof.

3. Analysis and optimality

In this section we introduce optimality criteria for an infinite time horizon with respect to $V^f(t)$ or $v^f(s)$ and investigate their relations with the Laurent series expansion for $v^f(s)$. We give conditions assuring the continuity of the terms of the Laurent series expansion for $v^f(s)$ and establish the existence of optimal policies.

A policy f is called *s-discount optimal* if $v_i^f(s) \geqslant v_i^h(s)$ for all $h \in F$, $i \in E$. There are quite weak assumptions for the existence of stationary s-discount optimal policies for fixed $s > 0$, cf. Harrison [17], Schäl [30], Van Nunen and Wessels [38] and Van Nunen [37]. However, in the case of unbounded rewards these assumptions require s to be fixed or bounded away from zero. From our analysis it will follow that assumptions 1, 2 and a continuity assumption guarantee for each $s > 0$ the existence of an s-discount optimal policy. We shall come back to this at the end of this section.

A criterion which considers the behavior of $v^f(s)$ for s small, was introduced for the discrete time, finite state and action case by Blackwell [2]. We shall generalize it to the denumerable state SMDC and discern two variants, thereby following Dekker and Hordijk [7].

DEFINITION 3.1

A policy f is called *Blackwell optimal* if there exists for every $i \in E$, $h \in F$ and $s(i, h) \in \mathbb{R}$ such that $v_i^f(s) \geqslant v_i^h(s)$, for all $0 < s \leqslant s(i, h)$. A policy f is called *strongly Blackwell optimal* if $s_0 \equiv \inf_{i \in E, h \in F} s(i, h) > 0$.

In the finite state and action case, Blackwell optimality is the same as strong Blackwell optimality. In that case a Blackwell optimal policy is s-discount optimal for all $0 < s \leqslant s_0$, for some $s_0 > 0$.

Another widely used criterion is average optimality. A policy f is called *average optimal* if $\liminf_{t \to \infty} t^{-1}[V_i^f(t) - V_i^h(t)] \geqslant 0$ for all $i \in E$ and $h \in F$. The existence of a Laurent series expansion for $v^f(s)$, the Laplace–Stieltjes transform of $V_i^f(t)$, implies that $g_i(f) = \lim_{t \to \infty} t^{-1} V_i^f(t)$ exists and is equal to $\lim_{s \downarrow 0} s v_i^f(s) = V_i^{(-1)}(f)$. We call $g_i(f)$ the *average reward* under policy f when the system starts in state i. From this we see that average optimality corresponds to optimality of the first term of the Laurent series expansion for $v^f(s)$.

For the average rewards only the asymptotic behaviour of the semi-Markov process counts, hence any reward earned in a finite period yields no contribution to it. This was the reason behind the development of more sensitive criteria. Veinott [40] defined a policy f to be *n-discount optimal* if $\lim_{s \downarrow 0} s^{-n}[v_i^f(s) - v_i^h(s)] \geqslant 0$ for all $i \in E$, $h \in F$. Recalling the Laurent series expansion for $v^f(s)$ we see that a policy f is n-discount optimal if it maximizes lexicographically the first $(n + 2)$ terms of the Laurent series expansion for all states simultaneously. It is not difficult to see that Blackwell optimality corresponds to lexicographic

optimality of all terms of the Laurent series expansion. We shall be able to maximize these terms uniformly in the initial state i once we have continuity of these terms in f. In the following, we call $x(f)$ a *vector function* on F if, for each $f \in F$, $x(f)$ is an element of V^μ. Elements of V^μ can also be considered as vectors in $\mathbb{R}^\infty$ since E is the set $\{1, 2, \ldots\}$. In the same way, we speak of *matrix functions A(f)* and denote the set of μ-bounded matrix functions by M^μ. In the sequel we will use the concept of μ-continuity which was introduced in Dekker and Hordijk [7]. It is defined as follows.

DEFINITION 3.2

(a) A vector function $x(f)$ and matrix function $A(f)$ on F are called pointwise continuous on F if for all $i, j \in E$, $f^{(0)} \in F$ and all sequences $f^{(n)} \to f^{(0)}$ in F we have

$$\lim_{f^{(n)} \to f^{(0)}} x_i(f^{(n)}) = x_i(f^{(0)}) \quad \text{and} \quad \lim_{f^{(n)} \to f^{(0)}} A_{ij}(f^{(n)}) = A_{ij}(f^{(0)}). \tag{3.1a}$$

(b) A matrix function $A(f)$ on F is called μ-continuous on F if for all $i \in E$, $f^{(0)} \in F$ and all sequences $f^{(n)} \to f^{(0)}$ we have

$$\lim_{f^{(n)} \to f^{(0)}} \sum_j \left| A_{ij}(f^{(n)}) - A_{ij}(f^{(0)}) \right| \mu_j = 0. \tag{3.1b}$$

The following lemma states the relations between μ-continuity and other forms of continuity. Recall that for any matrix function $A(f) \in M^\mu$ we have $\sum_j A_{ij}(f)\mu_j < \infty$.

LEMMA 3.1

For any matrix function $A(f) \in M^\mu$ the following assertions are equivalent:

(a) $A(f)$ is μ-continuous on F.

(b) Both $A(f)$ and $|A(f)|\mu$ are pointwise continuous.

(c) For any sequence $x^{(n)}$, $n = 1, 2, \ldots$, pointwise converging to $x^{(0)}$ with $\sup_{n=0,1,\ldots} \| x^{(n)} \|_\mu < \infty$ and any sequence $f^{(n)} \to f^{(0)}$, we have $A(f^{(n)})x^{(n)}$ converges pointwise to $A(f^{(0)})x^{(0)}$.

Proof

See Dekker [9].

μ-continuity has the advantage over pointwise continuity of being preserved when multiplied by other μ-continuous matrix functions as stated in the following lemma which proof is straightforward and therefore omitted.

LEMMA 3.2

If $A(f)$ and $B(f)$ are μ-continuous matrix functions on F, then

(a) $A(f) + B(f)$ is also μ-continuous,

(b) if $\| B(f) \|_\mu < c$ for all $f \in F$ and some $c \in \mathbb{R}$ then $A(f)B(f)$ is μ-continuous.

Concerning continuity we make the following assumption.

ASSUMPTION 5
(a) The matrix $D(f)$ is μ-continuous on F.
(b) The matrices $Q^{(k)}(f)$ are μ-continuous on F, for $k = 0, 1, 2, \ldots$.
(c) The vectors $R^{(k)}(f)$ are pointwise continuous on F, for $k = 0, 1, 2, \ldots$.

Recall that $Q^{(0)}(f) = P(f)$. Hence by assumption 5, lemma 3.2 and (2.19) it follows that $\Pi(f)$ is also μ-continuous. This enables us to establish the following lemma.

LEMMA 3.3

Suppose the vector functions $b(f)$, $c(f) \in V^\mu$ are pointwise continuous on F, then the μ-bounded solution $x(f)$ of eqs. (2.25) is also pointwise continuous on F.

Proof

The μ-bounded solution $x(f)$ is given by eq. (2.26). Since the denominator is bounded away from zero by (2.24), the quotient in (2.26) is pointwise continuous by assumption 5, the subsequent remark, lemma 3.2 and the assumptions of this lemma. The same arguments provide the pointwise continuity of the whole expression in (2.26). □

THEOREM 3.4

The vectors $V^{(k)}(f)$ in the Laurent series expansion for $v^f(s)$ are pointwise continuous on F.

Proof

From theorem 2.4 we know that $b^{(k)}(f)$, $c^{(k)}(f)$ and $V^{(k)}(f)$ are μ-bounded for all k. We shall establish the pointwise continuity of $b^{(k)}(f)$, $c^{(k)}(f)$ and $v^{(k)}(f)$ by induction in k.

For $k = -1$ we have $b^{(-1)}(f) = 0$ and $c^{(-1)}(f) = R^{(0)}(f)$ which is pointwise continuous by assumption 5(c). Lemma 3.3 provides the pointwise continuity of $V^{(-1)}(f)$.

Suppose $b^{(n)}(f)$, $c^{(n)}(f)$ and $V^{(n)}(f)$ are pointwise continuous for $n = -1, \ldots, k-1$. From expressions (2.29) and (2.30) we see that the pointwise continuity of $b^{(k)}(f)$ and $c^{(k)}(f)$, respectively, follows from assumption 5, lemma 3.2 and the induction hypothesis. Lemma 3.3 again provides pointwise continuity of $V^{(k)}(f)$ which completes the induction step. □

We shall now turn our attention to the optimization problem. The analysis proceeds in the same way as in Dekker and Hordijk [7]. In the sequel we use different reward functions $R(t)$ on E with one specific transition structure $Q^f(t)$. We therefore introduce an operator H^f which assigns the Laurent series expansion for its total discounted rewards $v^f(s)$, to each vector $R(t)$ satisfying assumption 5(c), with $Q^f(t)$ as transition structure. First we define the appropriate space of Laurent series. Let x^T stand for the transpose of vector x, the constant d is given by (2.27).

DEFINITION 3.3

$$LS^{\infty} \equiv \left\{ y = (y_1, y_2, \ldots)^T \mid y_i = \sum_{k=-1}^{\infty} a_i^{(k)} s^k, \right.$$

$$\left. a^{(k)} \in V^{\mu},\ i \in E \text{ and } \limsup_{k \to \infty} \left(\| a^{(k)} \|_{\mu} \right)^{1/k} \leqslant d^{-1} \right\}.$$

LS^{∞} consists of vectors with Laurent series as elements. All Laurent series have s^{-1} as leading term. It is a linear space with respect to termwise addition, subtraction and scalar multiplication. It is ordered in the following way. Let $y_i(s)$ be the value of the Laurent series y_i in s.

DEFINITION 3.4

For $y \in LS^{\infty}$, $y = (y_1, y_2, \ldots)^T$,

(a) $y_l \geqslant 0$ if $\liminf_{s \downarrow 0} y_i(s) s^{-n} \geqslant 0$, for all $n \geqslant -1$, $i \in E$,

(b) $y_l > 0$ if $y_l \geqslant 0$ and $\liminf_{s \downarrow 0} y_{i_0}(s) s^{-n_0} > 0$, for some $i_0 \in E$, $n_0 \geqslant -1$,

(c) $y_l \geqslant x$ if $y - x_l \geqslant 0$; $y_l = x$ if $y_l \geqslant x$ and $x_l \geqslant y$.

Remark

If $y_l = 0$ then $a_i^{(k)} = 0$ for $i \in E$ and $k = -1, 0, 1, \ldots$.

In order to prove the existence of s-discount optimal policies, we introduce the following operator on V^{μ}

$$B_s u = \max_{f \in F} \left[r^f(s) + q^f(s) u \right], \tag{3.2}$$

for any $u \in V^{\mu}$. The maximization over $f \in F$ breaks down into maximization per state over a from $A(i)$. By assumptions 3 and 5 and lemma 3.2 we have, in each state, a real-valued bounded and continuous function of the decision a, element of the compact set $A(i)$ and hence a maximum exists. B_s is called the *maximal one-step-reward operator with discount rate s*. It is not difficult to give conditions under which B_s is a contraction for that s. It then follows from the fixed point theorem that B_s has a unique fixed point, say $v(s)$. Moreover, $v_i(s)$ is equal to $\sup_{f \in F} v_i^f(s)$ and the policy which takes maximizing actions in (3.2) is s-discount

optimal. It is not possible to generalize this method for Blackwell optimality since we cannot make use of a fixed point theorem on LS^∞. Therefore we introduce a new operator $H^f(s)$ on V^μ for a fixed s and its generalization H^f on LS^∞. We shall set up the theory directly for LS^∞ and Blackwell optimality, but the entire analysis can also be carried out for a fixed s and s-discount optimality, if instead of the lexicographic ordering "$_l\geqslant$" on LS^∞, the normal ordering "$\geqslant$" on V^μ for the functions evaluated in that s is used. Let H^f be defined by

$$H^f y \equiv \sum_{n=0}^{\infty} [q^f]^n y, \tag{3.3}$$

for $y \in LS^\infty$. This does not always yield an element of LS^∞. In order to overcome this difficulty, we work with a slightly different operator:

$$(\overline{H}^f y)(s) = \sum_{n=0}^{\infty} [q^f(s)]^n [sy(s)]. \tag{3.4}$$

Since $sy(s)$ is a vector consisting of power series with a convergence radius of at least d and d is larger than $(c_3)^{-1}$, one can obtain $\overline{H}^f y$ by solving v from $v = sy(s) + q^f(s)v$. It follows from theorem 2.4 that there exists a Laurent series expansion for v which satisfies the requirements of LS^∞. Hence we can regard $\overline{H}^f y(s)$ as the expansion for the discounted rewards $v^f(s)$ where $sy(s)$ is the expansion of the transform of the appropriate reward vector.

When the immediate rewards are a positive vector (that is, at least one component is positive and the other components are nonnegative), the total expected s-discounted rewards are also a positive vector for every $s > 0$. Hence its Laurent series expansion is positive in the sense of our ordering. This important property is generalized in the following theorem.

THEOREM 3.5

For any $f \in F$, $\overline{H}^f$ is a positive operator on LS^∞, that is, $y \in LS^\infty$, $y_l > 0$ implies $\overline{H}^f y_l > 0$.

Proof

The proof follows similar lines as that of theorem 4.2 in Dekker and Hordijk [7]. The first part we prove is that for each state $i \in E$, $(\overline{H}^f y)_{i\,l} \geqslant 0$, i.e. the first nonzero coefficient (if it exists) in the Laurent series for $(\overline{H}^f y)_i$ is positive. This will be done by carefully checking how such coefficients originate. Subsequently, we shall prove that $(\overline{H}^f y)_{j\,l} > 0$ for some $j \in E$. For the sake of convenience we skip the dependence on f in the notation. Let $i \in E$ be fixed and $y_j = \sum_{k=-1}^{\infty} a_j^{(k)} s^k$.

Suppose i is contained in a minimal closed set, say C. All states in C are positive recurrent by the remarks following (2.22), hence we have $\Pi_{jh} > 0$ for all $j, h \in C$. Consider the power series y_j, $j \in C$. Let n be the index of the first

nonzero coefficient over all these power series (if such an index exists), i.e.

$a_j^{(k)} = 0$ for all $j \in C$ and $-1 \leqslant k < n$,

$a_j^{(n)} \geqslant 0$ for all $j \in C$ and $a_h^{(n)} > 0$ for some $h \in C$.

Having $r(s) = sy(s)$ as expansion for the transform of the reward vector and using the notation of theorem 2.4 one can easily see that for all $j \in C$, $c_j^{(k)} = 0$, $V_j^{(k)} = 0$ and $b_j^{(k+1)} = 0$, $-1 \leqslant k < n-1$ and also that $V_j^{(n-1)} > 0$ for all $j \in C$. Hence $(\overline{H}^f y)_{j\,l} > 0$ for all $j \in C$. If no such index exists, i.e. if all $y_{j\,l} = 0$, $j \in C$, we also get $V_j^{(k)} = 0$, $k = -1, 0, \ldots$ for all $j \in C$ and $(\overline{H}^f y)_{j\,l} = 0$, $j \in C$.

Suppose i is a transient state under P. Hence we have $V_i^{(k)} = (Db^{(k)})_i + z_i^{(k)}$, $k = -1, 0, \ldots$ with

$$z_i^{(k)} = \sum_j \Pi_{ij} \frac{\left[\Pi c^{(k)} - \Pi Q^{(1)} D b^{(k)}\right]_j}{[\Pi \tau]_j}.$$

Let $R(i) \equiv \{ j \in E \mid \Pi_{ij} > 0\}$ and $T(i) \equiv \{ j \in E \mid D_{ij} > 0 \text{ and } \Pi_{ij} = 0\}$, then $B(i) \equiv R(i) \cup T(i)$ consists of all states accessible from i, where $R(i)$ are the recurrent states and $T(i)$ the transient states. Note that $B(j) \subset B(i)$ for each $j \in B(i)$. In addition, note that for any $j \in B(i)$, $z_j^{(k)}$ is independent of the values of $b_h^{(k)}$ and $c_h^{(k)}$ for any $h \in T(i)$. Consider the power series y_j, $j \in B(i)$ and let n be the index of the first nonzero coefficient over all these power series (if such an index exists). For each $h \in B(i)$ which $a_h^{(n)} > 0$, we shall consider its contribution to $V_j^{(k)}$, $k = -1, 0, \ldots$, $j \in B(i)$. Since $a_j^{(k)} = 0$, $-1 \leqslant k < n$ for all $j \in B(i)$ it is easily seen that both $c_j^{(k)} = 0$, $b_j^{(k+1)} = 0$ and $V_j^{(k)} = 0$ for $-1 \leqslant k < n-1$, $j \in B(i)$. Furthermore, $c_j^{(n-1)} \geqslant 0$, $j \in B(i)$ and $c_h^{(n-1)} > 0$. If $h \in R(i)$ then $v_h^{(n-1)} > 0$ and also $V_i^{(n-1)} > 0$, since $\Pi_{ih} > 0$. If $h \in T(i)$ then $V_h^{(n-1)} = 0$; however, $b_h^{(n)} > 0$ and hence $V_i^{(n)} > 0$, since $D_{ih} > 0$. In both cases we have $(\overline{H}^f y)_{i\,l} > 0$.

The fact that $\overline{H}^f y_l > 0$ now remains to be proven. Since $y_l > 0$, there exists a state j such that $y_{j\,j} > 0$, say $a_j^{(k)}$, is the first nonzero coefficient. Since either $\Pi_{jj} > 0$ or $D_{jj} > 0$, we can see by the previous arguments that either $V_j^{(k+1)} > 0$ or $V_j^{(k)} > 0$, which completes the proof. □

H^f can be considered as the inverse operator of $[I - q^f]$, that is, for each $f \in LS^\infty$

$$H^f \left[I - q^f\right] y_l = y, \tag{3.5}$$

which follows directly from the definition of H^f, by taking s fixed and small enough. In the finite state space model $H^f(s)$ is, for a fixed $s > 0$, the inverse matrix of $[I - q^f(s)]$.

We shall now give the generalization B of B_s on LS^∞.

$$(By)(s) \equiv \operatorname*{lex.max}_{f \in F} \left\{ \sum_{k=-1}^{\infty} s^k \left[(-1)^k R^{(k)}(f) + \sum_{n=0}^{k+1} Q^{(n)}(f) a^{(k-n)} \right] \right\}, \tag{3.6}$$

for $y \in LS^{\infty}$, $y(s) = \sum_{k=-1}^{\infty} a^{(k)} s^k$ with $a^{(k)} \in V^{\mu}$. $R^{(k)}(f)$ and $Q^{(k)}(f)$ are defined in assumption 3, $R^{(-1)}(f) \equiv 0$, $f \in F$. "Lex.max" stands for lexicographic maximum, that is, we maximize the terms of the Laurent series lexicographically, which corresponds to maximizing in our order relation "$_l \geqslant$" on LS^{∞}. The maximization is taken componentwise, so we can restrict ourselves to $A(i)$ instead of F. Let $i \in E$ be fixed, $a = f(i)$ and let

$$x^{(k)}(a) \equiv (-1)^k R_i^{(k)}(f) + \sum_{n=0}^{k+1} (-s)^n \Sigma_j Q_{ij}^{(n)}(f) a_j^{(k-n)},$$

$$k = -1, 0, \ldots, a \in A(i).$$

If follows from assumptions 3 and 5 and lemma 3.2 that $x^{(k)}(a)$, $k = -1, 0, \ldots$ is a continuous and bounded function of a. Since $A(i)$ is compact, there exists a nonempty subset $A_{-1}(i) \subset A(i)$ in which all actions maximize $x^{(-1)}(a)$. Since $x^{(-1)}(a)$ is continuous in a, $A_{-1}(i)$ is closed and hence compact. Within $A_{-1}(i)$ there exists a second subset $A_0(i)$ consisting of actions maximizing $x^{(0)}(a)$. In this way we obtain a sequence of subsets $A_{-1}(i) \supset A_0(i) \supset A_1(i) \supset \ldots$, with each set closed and nonempty. Hence $A_{\infty}(i) \equiv \cap_{n=-1}^{\infty} A_n(i)$ is nonempty and any action of $A_{\infty}(i)$ maximizes the coefficients of $\sum_{k=-1}^{\infty} x^{(k)}(a) s^k$ lexicographically. B is thus a well-defined operator.

It is not possible to maximize $v^f(s)$ in the same way since the maximization cannot be done componentwise. Therefore, we take a weighted sum over all components. Let

$$w^f(s) \equiv \sum_{i=1}^{\infty} \frac{2^{-i}}{\mu_i} v_i^f(s), \quad f \in F. \tag{3.7}$$

$w^f(s)$ is a single Laurent series, say $w^f(s) = \sum_{k=-1}^{\infty} w^{(k)}(f) s^k$, with

$$w^{(k)}(f) = \sum_{i=1}^{\infty} \frac{2^{-i}}{\mu_i} V_i^{(k)}(f), \quad f \in F.$$

We have

$$|w^{(k)}(f)| \leqslant \|V^{(k)}(f)\|_{\mu}, \quad f \in F.$$

Since $V_i^{(k)}(f)/\mu_i$ is bounded in both i and f and each term is continuous in f, $w^{(k)}(f)$ is also continuous and bounded in f for each $k \geqslant -1$. Using a similar argumentation as for the existence of the operator B gives the following lemma.

LEMMA 3.6

There exists a policy $f_0 \in F$ that maximizes w^f lexicographically, i.e. $w^{f_0} {}_l \geqslant w^f$ for all $f \in F$.

This f_0 will appear to be Blackwell optimal.

THEOREM 3.7

Under assumptions 1,...,5 the following assertions hold for the Blackwell optimality equations $Bv_l = v$ in LS^∞, i.e.,

$$\operatorname{lex.max}_{f \in F} \left[r^f + q^f v \right]_l = v, \quad v \in LS^\infty; \tag{3.8}$$

(a) there exists a unique solution v_0 to it,
(b) a policy, say f_0, which maximizes w^f (cf. (3.7)) lexicographically, is Blackwell optimal and, moreover, $v_0 = v^{f_0}{}_l = \operatorname{lex.max}_{f \in F} v^f$
(c) any conserving policy f, i.e. for which $r^f + q^f v_0{}_l = v_0$ is Blackwell optimal.

Proof

The proof follows the same lines as that of theorems 3.2 and 3.3 of Dekker and Hordijk [7], but we will give it for completeness. We first show that v^{f_0} is a solution to (3.8), where f_0 is derived from lemma 3.6. From (2.10) we see that $r^{f_0} + q^{f_0} v^{f_0} - v^{f_0}{}_l = 0$ in LS^∞. Suppose there is a policy $\tilde{g}$ such that for some state $i \in E, (r^{\tilde{g}} + q^{\tilde{g}} v^{f_0} - v^{f_0})_{i\,l} > 0$, then define policy g by $g(i) = \tilde{g}(i)$ and $g(j) = f_0(j)$ for $j \neq i$. For this policy g it holds that $[r^g + (q^g - I) v^{f_0}]_{j\,l} \geqslant 0$ for all $j \neq i$ and ${}_l > 0$ for $j = i$. It then follows from theorem 3.5 that $s^{-1} \overline{H}^g [r^g + (q^g - I) v^{f_0}]_l > 0$. From (3.5) we see that this expression is equal to $v^g - v^{f_0}{}_l > 0$. However, this implies that $w^g - w^{f_0}{}_l > 0$, which contradicts lemma 3.6. Consequently, for all policies $f \in F$, we have $r^f + (q^f - I) v^{f_0}{}_l \leqslant 0$ and hence $Bv^{f_0}{}_l = 0$.

In the same way as for policy g it can be shown, for any policy $f \in F$, that $v^f - v^{f_0}{}_l \leqslant 0$ and hence $v^{f_0}{}_l = \operatorname{lex.max}_{f \in F} v^f$, which proves part (b).

Finally, suppose that we also have a solution $\tilde{v} \in LS^\infty$ to the Blackwell optimality equations and that policy $\tilde{f}$ takes maximizing actions for this solution, i.e. $r^{\tilde{f}} + q^{\tilde{f}} \tilde{v}_l = \tilde{v}$. From lemma 2.1 it follows that $\tilde{v}_l = v^{\tilde{f}}$. Since $B\tilde{v}_l = \tilde{v}$ it holds that $r^{f_0} + (q^{f_0} - I) \tilde{v}_l \leqslant 0$. Using $\overline{H}^{f_0}$ we obtain that $v^{f_0}{}_l \leqslant \tilde{v}$. Hence it follows from part (b) that $v^{f_0}{}_l = \tilde{v}$, which shows both parts (a) and (c). □

Let $H^f(s)$ denote H^f evaluated for a fixed s, i.e., $H^f(s) = \sum_{n=0}^\infty [q^f(s)]^n$. $H^f(s)$ is a nonnegative matrix with positive diagonal elements, which implies that it is a positive operator for every $s > 0$. It is not difficult to see that we can restate lemma 3.6 and theorem 3.7 for s-discount optimality if we evaluate all functions and operators in s and use the normal " $\geqslant$ " ordering. In fact, the only assumptions needed for applying these theorems for a fixed s, apart from assumptions 1 and 2, are the μ-continuity and pointwise continuity of $\sum_{n=0}^\infty [q^f(s)]^n$ and $r^f(s)$, respectively. If $q^f(s)$ is μ-continuous, then a contraction property of $q^f(s)$, i.e. $\| q^f(s) \|_\mu \leqslant c < 1$, for all $f \in F$, immediately implies the μ-continuity and the μ-boundedness of $\sum_{n=0}^\infty [q^f(s)]^n$. In this way we can easily establish the existence of a unique μ-bounded solution to the discount optimality equations with discount rate s. Any conserving policy is s-discount optimal and its total discounted reward is equal to the solution of the optimality equations.

It has already been stated that Blackwell optimality implies n-discount optimality for all n. It follows quite directly from our analysis that we can also define n-discount optimality equations.

4. Single server queueing network

In this section we consider a network with service facilities $1, \ldots, K$. The state space E is $\mathbb{N}_0^k$ and consists of all k-tuples, with i_k the number of jobs at facility of queue k, $1 \leqslant k \leqslant K$. The service time distribution of a job at queue k is denoted by F_k. After completing service at facility k, a departing job joins the queue at facility l with probability r_{kl}. The stochastic network is supposed to be open, so the routing matrix $R = (r_{kl})$ is assumed to be transient. The arrival processes to the network are independent Poisson processes. The rate of arrivals to facility k is λ_k. Let $\lambda = \sum_k \lambda_k$. The throughputs at the facilities can be found as the unique solution of the traffic equations,

$$\gamma_j = \sum_{i=1}^{K} \gamma_i r_{ij} + \lambda_j .$$

As in Klimov [22] we assume the following ergodicity condition:

$$\sum_{i=1}^{K} \gamma_i \beta_i < 1 ,$$

where β_i is the first moment of the services time distribution F_i. There is only one server in the network. At the completion of a service the decision maker has to decide which of the nonempty queues has to be served next. The optimal control of the server can be formulated as a semi-Markov decision chain. The decision epochs are the service completion times together with the arrival times to an empty network. So $T_{n+1} - T_n$ has distribution F_k resp. E^λ (E^λ denotes the nonnegative exponential distribution with mean λ^{-1}) if at decision epoch T_n the server starts serving queue k resp. the system is empty. The state S_n at T_n is the state of the network just after a service completion resp. just after the arrival of a job to an empty network. Action or decision k means that the server will serve queue k next. We only allow non-idling policies, which means $A(i_1, \ldots, i_K) = \{ a \mid i_a > 0, 1 \leqslant a \leqslant K \}$. In the transition probabilities $Q_{ij}^a(x)$ as defined in (2.1) we use the notation $i = (i_1, \ldots, i_K)$ and $k = (k_1, \ldots, k_K)$ with $i_l, k_l \geqslant 0$, $l = 1, \ldots, K$. The symbols 0 resp. e_a denote the vector with all components equal to zero resp. the vector for which component a is the only nonzero component and equal to one.

$$Q_{0k}^a(x) = \lambda_j \lambda^{-1}(1 - e^{-\lambda x}) \quad \text{when } k_i = 0,\ i \neq j \text{ and } k_j = 1;$$

$$Q^a_{i(i+k)}(x) = \int_{t=0}^{x} \prod_j \left(\frac{e^{-\lambda_j t}(\lambda_j t)^{k_j}}{k_j!} \right)$$

$$\times \left\{ r_{aa} + \frac{\lambda_a t}{k_a + 1}\left(1 - \sum_j r_{aj}\right) + \sum_{j \neq a} \frac{\lambda_a}{\lambda_j} \frac{k_j}{k_a + 1} r_{aj} \right\} \mathrm{d}F_a(t);$$

$$Q^a_{i(i+k-e_a)}(x) = \int_{t=0}^{x} \prod_{j \neq a} \left(\frac{e^{-\lambda_j t}(\lambda_j t)^{k_j}}{k_j!} \right) e^{-\lambda_a t}\left(1 - \sum_j r_{aj}\right) \mathrm{d}F_a(t)$$

with $k_a = 0$,

and the other transition probabilities are zero. We recall that $R^a_i(x)$ denotes the expectation of the income during the time interval $[0, \min(x, T_1)]$, given $S_0 = i$ and $D(i) = a$. Let us assume that the cost structure consists of a nondecreasing holding cost rate $h(i) \geqslant 0$ when i jobs are in the network, a service cost rate $s(a) \geqslant 0$ when queue a is served and an instantaneous cost $c(i, a)$. The expression for the expected holding cost during $[0, \min(x, T_1)]$ is messy. However, it is easy to see that the following expression is an upper bound:

$$\left(\sum_j Q^a_{ij}(\infty) h(j + e_a) \right) \left(\int_0^x t \, \mathrm{d}F_a(t) \right).$$

Let us for the moment assume that this summation is finite. Hence,

$$(R^a_i)^+(x) = c^-(i, a),$$

$$(R^a_i)^-(x) \leqslant \left(\sum_j Q^a_{ij}(\infty) h(j + e_a) + s(a) \right) \left(\int_0^x t \, \mathrm{d}F_a(t) \right) + c^+(i, a).$$

It is easily verified that assumption 1 holds. For the service time at facility k we assume

$$F_k(t_1) \geqslant c_1, \tag{4.1}$$

for some positive t_1 and c_1, and

$$1 - F_k(t) \leqslant c_2 \, e^{-s_1 t}, \tag{4.2}$$

for some positive $s_1 < \lambda$ and $c_2 > 1$ and all $t \geqslant 0$. In our opinion this assumption will always be satisfied by distributions encountered in practice. However, in theory it is restrictive. If only average optimality is considered the condition can be relaxed. Since there are only K queues there is no restriction in assuming that the inequalities hold for all $k = 1, \ldots, K$ simultaneously. For our bounding vector μ we take

$$\mu_{(i_1, \ldots, i_k)} = \begin{cases} \mu_0 & i_1 = \ldots = i_k = 0, \\ (1 + x_1)^{i_1} \ldots (1 + x_k)^{i_k} & \text{otherwise}, \end{cases}$$

where $\mu_0 \geqslant 1$ and $x_1, \ldots, x_k > 0$. We assume that $\| h(\cdot) \|_\mu < \infty$ and $\| c(\cdot) \|_\mu < \infty$ with $c(i) = \max_a c(i, a)$. Remark that this is the case if $h(i)$ and $c(i)$ are bounded by a multinomium in $(i_1, \ldots, i_k)$. Assuming relation (4.2) it is shown in section 3.3 of Spieksma [35] that for some $\beta < 1$,

$$\sum_{j \neq 0} P_{ij}(f)\mu_j \leqslant \beta\mu_i, \quad i \in E,\ f \in F \tag{4.3}$$

with $x_1, \ldots, x_k$ sufficiently close to zero and μ_0 sufficiently large. The relation (4.3) implies the μ-geometric recurrence property of Dekker and Hordijk [8]. It follows from their lemma 3.2 that for some $c < \infty$ and all $f \in F$,

$$\| P^k(f) \|_\mu < c, \quad k \in \mathbb{N}_0. \tag{4.4}$$

Define,

$$F(x) = \min\{ E^\lambda(x),\ F_k(x),\ k = 1, \ldots, K \},$$

and let $N(x)$ be the number of renewals with lifetime distribution $F(x)$. Clearly,

$$\sum_j M_{ij}^f(t)\mu_j \leqslant \sum_{k=0}^{\infty} P(N(t) = k) \sum_j P_{ij}^k(f)\mu_j.$$

With (4.4) we find that for some $c < \infty$,

$$\| M^f(t) \|_\mu \leqslant \frac{t}{\tau} + c,$$

with τ the first moment of F. Note that (4.1) implies $\tau > 0$. Hence assumption 2(b) is satisfied. Recall that e is the vector with all components equal to one. Since $\mu_{j+e} \leqslant c_3\mu_j$ for some constant c_3 we have,

$$\sum_j Q_{ij}^f(\infty) h(j + e_{f(i)}) \leqslant \sum_j P_{ij}(f) h(j + e) \leqslant c_3 \| h(\cdot) \|_\mu \| P(f) \|_\mu \mu_i < \infty,$$

and assumption 2(a) holds as well. Define

$$G(x) = \max\{ E^\lambda(x),\ F_k(x),\ k = 1, \ldots, K \},$$

then

$$1 - G(x) \leqslant c_2\, \mathrm{e}^{-s_1 x},$$

and hence, for some constant c,

$$\int_0^\infty \frac{x^k}{k!}\, \mathrm{d}G(x) \leqslant c \int_0^\infty \frac{x^k}{k!}\, \mathrm{d}(1 - \mathrm{e}^{-s_1 x}) \leqslant c\left(\frac{1}{s_1}\right)^k.$$

Consequently,

$$\sum_j \int_0^\infty \frac{x^k}{k!}\, \mathrm{d}Q_{ij}^f(x)\mu_j \leqslant \int_0^\infty \frac{x^k}{k!}\, \mathrm{d}G(x) \sum_j P_{ij}(f)\mu_{j+e}$$

and

$$\| Q^{(k)}(f) \|_\mu \leqslant c_3 \| P(f) \|_\mu c\left(\frac{1}{s_1}\right)^k.$$

Hence we conclude that assumption 3(a) is satisfied. The verification of assumption 3(b) is done in a similar way. Assumption 4 follows from theorem 3.5 (see also the proof of corollary 3.6) of Dekker and Hordijk [8]. To verify assumption 5(b) it is, according to lemma 3.1, sufficient to show that $Q^{(k)}(f)$ and $Q^{(k)}(f)\mu$, $k = 1, 2, \ldots$ are pointwise continuous in f. Now suppose $f_n \to f$ as n tends to ∞, then for fixed $i \in E$, since $A(i)$ is finite, $f_n(i) = f(i)$ for n sufficiently large. Say $f_n(i) = f(i) = a$ for $n \geqslant n_0$. Hence for every k

$$Q_{ij}^{(k)}(f_n) = Q_{ij}^{(k)}(f) = \int_0^\infty \frac{x^k}{k!}\, \mathrm{d}Q_{ij}^{a}(x),$$

when $n \geqslant n_0$.

Similarly, the pointwise continuity of $Q^{(k)}(f)\mu$ and $R^{(k)}(f)$ in f follows from the finiteness of the action sets and therefore, assumptions 5(b) and 5(c) hold. Also the pointwise continuity of $P(f)\mu$ follows and assumption 5(a) is the assertion (ii) of theorem 3.10 in Dekker and Hordijk [8]. We conclude that theorem 3.7 is valid for the single server queueing network. The Blackwell optimality equations have a unique μ-bounded solution which equals the Laurent series expansion of a Blackwell optimal policy. In Klimov [22] the optimal order of service is derived for minimum average costs. If this optimal service order is unique then it also provides a Blackwell optimal policy.

References

[1] J.A. Bather, Optimal decision procedures for finite Markov chains, part I: Examples, Adv. Appl. Prob. 5 (1973) 328–339; part II: Communicating systems, Adv. Appl. Prob. 5 (1973) 521–540; part III: General convex systems. Adv. Appl. Prob. 5 (1973) 541–553.

[2] D. Blackwell, Discrete dynamic programming, Ann. Math. Statist. 33 (1962) 719–726.

[3] D. Blackwell, Discounted dynamic programming, Ann. Math. Statist. 36 (1965) 226–235.

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

[5] E. Cinlar, *Introduction to Stochastic Processes* (Prentice–Hall, Englewood Cliffs, NJ, 1975).

[6] J.S. De Cani, A dynamic programming algorithm for embedded Markov chains when the planning horizon is infinitely, Mgmt. Sci. 10 (1964) 716–733.

[7] R. Dekker and A. Hordijk, Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards, Math. Oper. Res. 13 (1988) 395–421.

[8] R. Dekker and A. Hordijk, Recurrence conditions for average and Blackwell optimality in denumerable state Markov decision chains, Technical report, Dept. of Math. and Comp. Sci, Univ. of Leiden (1989) to appear in Math. Oper. Res.

[9] R. Dekker, Denumerable Markov decision chains: optimal policies for small interest rates, Ph.D. thesis, Univ. of Leiden (1985).

[10] E.V. Denardo and B.L. Miller, An optimality condition for discrete dynamic programming with no discounting, Ann. Math. Statist. 39 (1968) 1220–1227.

[11] E.V. Denardo, Markov renewal programming with small interest rates, Ann. Math. Statist. 42 (1971) 477–496.

[12] H. Deppe, On the existence of average optimal policies in semiregenerative decision models, Math. Oper. Res. 9 (1984) 558–575.

[13] C. Derman, Denumerable state Markovian decision processes – average cost criterion, Ann. Math. Statist. 42 (1966) 1545–1553.
[14] A. Federgruen, A. Hordijk and H.C. Tijms, Denumerable state semi-Markov decision processes with unbounded costs, average cost criterion, Stoch. Proc. Appl. 9 (1979) 223–235.
[15] A. Federgruen, P.J. Schweitzer and H.C. Tijms, Denumerable undiscounted semi-Markov decision processes with unbounded rewards, Math. Oper. Res. 8 (1983) 298–313.
[16] A. Federgruen and H.C. Tijms, The optimality equation in average cost denumerable state semi-Markov decision problems, recurrency conditions and algorithms, J. Appl. Prob. 15 (1978) 356–373.
[17] J.M. Harrison, Discrete dynamic programming with unbounded rewards, Ann. Math. Statist. 43 (1972) 636–644.
[18] A. Hordijk, *Dynamic Programming and Markov Potential Theory*, Mathematical Centre Tract no. 51, Amsterdam (1974).
[19] A. Hordijk and K. Sladky, Sensitive optimality criteria in countable state dynamic programming, Math. Oper. Res. 2 (1977) 1–14.
[20] R.A. Howard, *Semi-Markovian Decision Processes*, Proc. Int. Statist. Inst., Ottawa, Canada (1963).
[21] W.S. Jewell, Markov-renewal programming I: formulation, finite return models; Markov-renewal programming II: infinite return models, example, Oper. Res. 11 (1963) 938–971.
[22] G.P. Klimov, Time-sharing service systems I, Th. Prob. Appl. 19 (1974) 532–551.
[23] J.B. Lasserre, Conditions for existence of average and Blackwell optimal stationary policies in denumerable Markov decision processes, J. Math. Anal. Appl. 136 (1988) 479–490.
[24] E. Mann, Optimality equations and sensitive optimality in bounded Markov decision processes, Optimization 16 (1985) 767–781.
[25] B.L. Miller, Finite state continuous time Markov decision processes with infinite planning horizon, J. Math. Anal. Appl. 22 (1968) 552–569.
[26] B.L. Miller and A.F. Veinott Jr., Discrete dynamic programming with a small interest rate, Ann. Math. Statist. 40 (1969) 366–370.
[27] S.M. Ross, Non-discounted denumerable Markovian decision models, Ann. Math. Statist. 39 (1968) 412–423.
[28] S.M. Ross, *Applied Probability Models with Optimization Applications* (Holden–Day, San Fransisco, 1970).
[29] H.L. Royden, *Real Analysis* (MacMillan, New-York, 2nd ed. 1968).
[30] M. Schäl, Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal, Z. Wahr. verw. Gebiete 32 (1975) 179–196.
[31] M. Schäl, On the second optimality equation for semi-Markov decision models, Technical report, Inst. Angew. Math., Univ. Bonn (1989) submitted for publication.
[32] P.J. Schweitzer, Perturbation theory and Markovian decision processes, Ph.D. dissertation, Mass. Inst. of Techn. (1965).
[33] L.J. Sennott, Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs, Oper. Res. 37 (1989) 626–633.
[34] L.J. Sennott, Average cost semi-Markov decision processes and the control of queueing systems, Prob. Eng. Inform. Sci. 2 (1989) 247–272.
[35] F.M. Spieksma, The existence of sensitive optimal policies in two multi-dimensional queueing models, this volume, pp. 273–296.
[36] Sh. Stidham, Jr. and R.R. Weber, Monotonic and insensitive optimal policies for control of queues with undiscounted costs, Oper. Res. 87 (1989) 611–625.
[37] J.A.E.E. van Nunen, *Contracting Markov Decision Processes*, Mathematical Centre Tract no. 71, Amsterdam (1976).
[38] J.A.E.E. van Nunen and J. Wessels, Markov decision processes with unbounded rewards, in:

Markov Decision Theory, eds. H.C. Tijms and J. Wessels, Mathematical Centre Tract no. 93, Amsterdam (1976).

[39] R.R. Weber and Sh. Stidham, Jr., Optimal control of service rates in networks of queues, Adv. Appl. Prob. 19 (1987) 202–218.

[40] A.F. Veinott Jr., On discrete dynamic programming with sensitive discount optimality criteria, Ann. Math. Stat. 40 (1969) 1635–1660.

[41] W.H.M. Zijm, The optimality equations in multichain denumerable state Markov decision processes with the average cost criterion: the bounded cost case, Stat. Decisions 3 (1985) 143–165.

[42] W.H.M. Zijm, The optimality equations in multichain denumerable state Markov decision processes with the average cost criterion: the unbounded cost case, C.Q.M. Note 22, Centre for Quant. Methods, Philips B.V. Eindhoven (1984).