stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 55/75 NOVEMBER

H.C. TIJMS

ON DYNAMIC PROGRAMMING WITH ARBITRARY STATE SPACE, COMPACT ACTION SPACE AND THE AVERAGE RETURN AS CRITERION

Prepublication

2e boerhaavestraat 49 amsterdam

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

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

AMS(MOS) subject classification scheme (1970): 90C40

On Dynamic Programming with Arbitrary State Space, Compact Action Space and the Average Return as Criterion

by

H.C. Tijms

ABSTRACT

A dynamic programming model with an arbitrary state space and a compact action space is considered. Under certain conditions it will be shown that there is an average return optimal stationary policy and that the optimality equation for the average return criterion applies. Besides some continuity assumptions on the immediate return and the transition probabilities, these conditions include the assumption that the transition probability functions associated with the stationary policies satisfy a recurrency condition known as the Doeblin condition in Markov chain theory. Finally, it will be proved that the value iteration method exhibits exponential convergence under certain conditions.

KEY WORDS & PHRASES: *Dynamic programming, arbitrary state space, compact action space, average return, optimal stationary policy, optimality equation, value iteration.*

1. INTRODUCTION

We consider a dynamic programming model specified by objects S, A, q and r, where S is a locally compact[*)] Borel subset of a complete, separable metric space, A is a compact metric space, q associates with each pair $(s,a) \in S \times A$ a probability distribution $q(\cdot|s,a)$ on the class $\mathcal{B}$ of Borel subsets of S, and r is a bounded Borel measurable function on $S \times A$. Let M be any constant such that $|r(s,a)| \leq M$ for all s and a. The set S denotes the state space of some system and A denotes the set of possible actions. The state of the system is observed at times $t = 0,1,\ldots$. When the system is in state s and action a is chosen, an immediate return $r(s,a)$ is received and the state next visited by the system is distributed according to $q(\cdot|s,a)$.

A policy π is a rule that prescribes for each time t which action to choose at time t as a Borel measurable function of the history $(s_1,a_1,\ldots,s_t)$ of the system up to time t. Denote by F the class of all Borel measurable functions $f\colon S \to A$. For any $f \in F$, let $f^{(\infty)}$ be the policy that chooses action $f(s)$ whenever the system is in state s. Such policies are called stationary policies.

For any policy π, denote by $\{X_t, t=0,1,\ldots\}$ and $\{a_t, t=0,1,\ldots\}$ the sequence of states and actions, and define

$$g(\pi,s) = \limsup_{t\to\infty} \frac{1}{t} E_\pi\{ \sum_{k=0}^{t} r(X_k,a_k) \mid X_0 = s\} \qquad \text{for } s \in S,$$

where E_π denotes the expectation given that policy π is used. A policy π^* is said to be average return optimal when $g(\pi^*,s) \geq g(\pi,s)$ for all π and $s \in S$.

For the dynamic programming model with a continuous state space and the average return as criterion optimality results were found in LIPPMAN [7], ROSS [9], WIJNGAARD [14] and YAMADA [15] (see HORDIJK [4], LIPPMAN [6] and ROSS [10] for the case of a denumerable state space). In this paper we shall give conditions under which an average return optimal stationary policy exists and the optimality equation for the average return criterion applies. Besides some continuity conditions on q and r, we shall impose on the Markov chains $\{X_t\}$ associated with the stationary policies a recurrency condition of the type known as the Doeblin condition in Markov chain theory

[*)] If A is finite the results of this paper also hold when the local compactness of S is not assumed (cf. the proof of Lemma 2 below).

(cf. DOOB [2]). Conditions of this type were also introduced in HORDIJK [4], WIJNGAARD [14] and YAMADA [15]. To obtain the optimality results for the average return criterion, we shall show that the family of functions given by the maximal discounted return function relative to some fixed state is bounded and equicontinuous and next follow the analysis in ROSS [9]. This will be done in section 2. Using the obtained result that the optimality equation for the average return criterion has a fixed point, it will be next shown in section 3 that the value iteration method exhibits exponential convergence under certain conditions.

2. OPTIMALITY RESULTS FOR THE AVERAGE RETURN CRITERION

We first introduce some notation. For any $f \in F$, let $q_f^n(B|s)$ with $s \in S$ and $B \in \mathcal{B}$ be the n-step transition probability function of the Markov chain $\{X_t\}$ associated with the stationary policy $f^{(\infty)}$. Also, let $q_f^0(B|s)$ be equal to 1 if $s \in B$, and 0 if $s \notin B$. For any signed measure μ on $(S,\mathcal{B})$, denote by the measure $|\mu|_v$ the total variation of μ. For any bounded Borel measurable function h on S, we have

$$(1) \qquad \left|\int_E h d\mu\right| \le N|\mu|_v(E) \qquad \text{for } E \in \mathcal{B}$$

when $|h| \le N$. Also, for any $E \in \mathcal{B}$,

$$(2) \qquad |\mu(E)| \le |\mu|_v(E) \le 2 \sup_{V \subseteq E} |\mu(V)|.$$

We now introduce the following assumptions A1-A5 under which it will be shown that the optimality equation for the average return criterion applies and there is an average return optimal stationary policy *).

A1. The function $r(\cdot,\cdot)$ is continuous on $S \times A$.

A2. $q(\cdot|s_n,a_n)$ converges weakly to $q(\cdot|s,a)$ as $s_n \to s$ and $a_n \to a$.

A3. (i) For each $s \in S$, the measure $q(\cdot|s,a_n)$ converges setwise to the measure $q(\cdot|s,a)$ as $a_n \to a$, or (ii) S is compact.

*) cf. also section 3 where we shall replace A5 by a different but related assumption under which the optimality results also hold.

A4. For each $s \in S$, $\sup_{B \in \mathcal{B}} |q(B|s',a) - q(B|s,a)| \to 0$ as $s' \to s$, uniformly in $a \in A$.

A5. There is an integer $\nu \geq 1$, numbers $\gamma > 0$ and $\rho > 0$ and, for each $f \in F$, a σ-finite measure ϕ_f on $(S,\mathcal{B})$ and a set $C_f \in \mathcal{B}$ such that

$$\phi_f(C_f) \geq \gamma \quad \text{and} \quad p_f^\nu(s'|s) \geq \rho \qquad \text{for all } s \in S \text{ and } s' \in C_f,$$

where $p_f^\nu(\cdot|s)$ denotes the density of the absolutely continuous component of $q_f^\nu(\cdot|s)$ with respect to ϕ_f.

We note that A4 actually states that, for each $s \in S$,

$$|Q(s',s,a)|_v(S) \to 0 \text{ as } s' \to s, \text{ uniformly in } a \in A, \tag{3}$$

where $|Q(s',s,a)|_v$ denotes the total variation of the signed measure $q(\cdot|s',a) - q(\cdot|s,a)$, cf. also p. 282 in SCHÄL [12].

A5 is an immediate extension of a condition introduced on p.197 in DOOB [2] and implies that, for each $f \in F$, the Markov chain $\{X_t\}$ associated with policy $f^{(\infty)}$ satisfies the so-called Doeblin condition. We observe that A5 covers the case in which there is an integer $\nu \geq 1$, a number $\rho > 0$ and, for each $f \in F$, a state s_f such that $q_f^\nu(\{s_f\}|s) \geq \rho$ for all $s \in S$ (let $\phi_f(B)$ be equal to 1 if $s_f \in B$, and 0 if $s_f \notin B$, and let $C_f = \{s_f\}$).

By the proof given on pp.197-198 in DOOB [2], we have under A5 that, for each $f \in F$, the Markov chain $\{X_t\}$ associated with policy $f^{(\infty)}$ has a (unique) stationary probability distribution $\pi_f(\cdot)$ such that, for all $s \in S$ and $B \in \mathcal{B}$,

$$|q_f^n(B|s) - \pi_f(B)| \leq (1-\rho\gamma)^{[n/\nu]} \qquad \text{for all } n \geq 1, \tag{4}$$

where $[x]$ denotes the largest integer less than or equal to x.

To derive the optimality results for the average return criterion, we first consider the discounted return model and introduce the following functions. Let $0 < \alpha < 1$. For any policy π, define

$$V_\alpha(\pi,s) = E_\pi\{\sum_{n=0}^{\infty} \alpha^n r(X_n,a_n) \mid X_0 = s\} \quad \text{for } s \in S,$$

and let $V_\alpha(s) = \sup_\pi V_\alpha(\pi,s)$, $s \in S$. A policy π^* is said to be α-optimal if $V_\alpha(\pi^*,s) = V_\alpha(s)$ for all $s \in S$. In MAITRA [8] the following results were proved (cf. also BLACKWELL [1] and SCHÄL [12]).

THEOREM 1. *Suppose that* A1-A2 *hold. Let* $0 < \alpha < 1$. *Then,* $V_\alpha(s)$, $s \in S$ *is the unique bounded continuous function satisfying*

$$(5) \qquad V_\alpha(s) = \max_{a\in A} \{r(s,a) + \alpha \int_S V_\alpha(s')q(ds'|s,a)\} \qquad \textit{for all } s \in S.$$

Furthermore, there exists an α*-optimal stationary policy* $f_\alpha^{(\infty)}$ *such that* $f_\alpha(s)$ *attains the maximum in the right side of* (5) *for all* $s \in S$.

In the following lemma we shall essentially use A5.

LEMMA 1. *Suppose that* A1-A2 *and* A5 *hold. Then,* $|V_\alpha(f^{(\infty)},s) - V_\alpha(f^{(\infty)},s')| \leq$ $\leq 4M\nu/\rho\gamma$ *for all* $0 < \alpha < 1$, $f \in F$ *and* $s,s' \in S$.

PROOF. We have

$$|V_\alpha(f^{(\infty)},s)-V_\alpha(f^{(\infty)},s')| \leq \sum_{n=0}^{\infty} \alpha^n \Big|\int_S r(y,f(y))\{q_f^n(dy|s)-q_f^n(dy|s')\}\Big|.$$

Now, by (2) and (4), the total variation of the signed measure $q_f^n(\cdot|s)$ + $- q_f^n(\cdot|s')$ is bounded by $4(1-\rho\gamma)^{[n/\nu]}$ for all $n \geq 1$, $f \in F$ and $s,s' \in S$. Next, using (1), we get the desired result. □

Following ROSS [9], fix some state s^* and, for any $0 < \alpha < 1$, define

$$v_\alpha(s) = V_\alpha(s) - V_\alpha(s^*) \qquad \text{for } s \in S.$$

LEMMA 2. *Suppose that* A1-A2 *and* A4-A5 *hold. Then the family* $\{v_\alpha(\cdot), 0<\alpha<1\}$ *of functions is bounded and equicontinuous on* S.

PROOF. Let $f_\alpha \in F$ be as in Theorem 1. Since $V_\alpha(s) = V_\alpha(f_\alpha^{(\infty)},s)$ for all s, it follows from Lemma 1 that $|v_\alpha(s)| \leq 4M\nu/\rho\gamma$ for all $0 < \alpha < 1$ and $s \in S$. Choose now s_0, s_1 and α, and assume that $v_\alpha(s_0) \geq v_\alpha(s_1)$. Put for abbreviation $\Delta = r(s_0,f_\alpha(s_0)) - r(s_1,f_\alpha(s_0))$. Then, using Theorem 1 and (1),

$$\begin{aligned} v_\alpha(s_0) - v_\alpha(s_1) &= V_\alpha(s_0) - V_\alpha(s_1) \le \\ &\le \Delta + \alpha \int_S V_\alpha(y)q(dy|s_0,f_\alpha(s_0)) - \alpha \int_S V_\alpha(y)q(dy|s_1,f_\alpha(s_0)) = \\ &= \Delta + \alpha \int_S v_\alpha(y)\{q(dy|s_0,f_\alpha(s_0)) - q(dy|s_1,f_\alpha(s_0))\} \le \\ &\le \Delta + (4M\nu/\rho\gamma)|Q(s_0,s_1,f_\alpha(s_0))|_v(S). \end{aligned}$$

It now follows that, for all $s,s' \in S$ and $0 < \alpha < 1$,

$$|v_\alpha(s')-v_\alpha(s)| \le \sup_{a\in A}|r(s',a)-r(s,a)| + (4M\nu/\rho\gamma) \sup_{a\in A}|Q(s',s,a)|_v(S).$$

Fix $s \in S$. Since S is locally compact, there is an open set 0 containing s such that $\bar{0}$ is compact. Then $r(\cdot,\cdot)$ is uniformly continuous on $\bar{0} \times A$. Now the above inequality and (3) imply that $\{v_\alpha(\cdot),\ 0<\alpha<1\}$ is equicontinuous at s. □

We now prove the optimality results for the average return criterion.

THEOREM 2. *Suppose that* A1-A5 *hold. Then there is a constant* g *and a bounded continuous function* $v(\cdot)$ *on* S *satisfying the optimality equation*

$$(6)\qquad g + v(s) = \max_{a\in A}\{r(s,a) + \int_S v(y)q(dy|s,a)\} \qquad \text{for all } s \in S.$$

Also, there is an average return optimal stationary policy $f^{(\infty)}$ *such that* $f(s)$ *attains the maximum in the right side of* (6) *for all* $s \in S$ *and, moreover,* $g(f^{(\infty)},s) = g$ *for all* $s \in S$.

PROOF. The proof of this theorem follows ROSS [9]. However, since in ROSS [9] the action space is finite, the proof of Theorem 2 of ROSS [9] needs some modifications. Using the Lemmas 1 and 2 and the fact that $|(1-\alpha)V_\alpha(s^*)| \le M$ for all $0 < \alpha < 1$, it follows from the Ascoli Theorem (e.g. ROYDEN [11]) that there is a constant g, a bounded continuous function $v(\cdot)$ on S and a sequence $\{\alpha_k\}$ with $\alpha_k \to 1$ as $k \to \infty$ such that $(1-\alpha_k)V_{\alpha_k}(s^*)$ converges to g as $k \to \infty$ and $v_{\alpha_k}(s)$ converges to $v(s)$ as $k \to \infty$ for all $s \in S$. Suppose now that we have proved (6) with sup instead of max. Then, by invoking the Lemmas 3.4, 4.1 and the Selection Theorem in MAITRA [8], it follows that there is a $f \in F$ such that $f(s)$ maximizes the right side of (6) for all $s \in S$. Next, the proof of Theorem 1 of ROSS [9] shows that policy $f^{(\infty)}$ is average return optimal and that $g(f^{(\infty)},s) = g$ for all s. It remains to prove (6). To do this, we distinguish between the cases (a) and (b).

Case (a). Part (i) of A3 holds. Fix $s_0 \in S$. Since A is a compact metric space, we can find an action $a_0 \in A$ and a subsequence $\{\beta_k\}$ of $\{\alpha_k\}$ with $\beta_k \to 1$ as $k \to \infty$ such that $f_{\beta_k}(s_0) \to a_0$ as $k \to \infty$. Then, by using part (i) of A3 and Proposition 18 on p.232 in ROYDEN [11],

$$\lim_{k\to\infty}\{r(s_0,f_{\beta_k}(s_0)) + \beta_k \int_S v_{\beta_k}(y)q(dy|s_0,f_{\beta_k}(s_0))\} =$$

$$= r(s_0,a_0) + \int_S v(y)q(dy|s_0,a_0).$$

Now, subtract $V_{\beta_k}(s^*)$ from both sides of (5) with $\alpha = \beta_k$ and $s = s_0$. Next, by letting $k \to \infty$, we get (6) for $s = s_0$ with a_0 as maximizing action.

Case (b). Part (ii) of A3 holds. Since S is compact, it follows from the Ascoli Theorem that the convergence of $v_{\alpha_k}(\cdot)$ to $v(\cdot)$ as $k \to \infty$ is uniform on S. Now, subtract $V_{\alpha_k}(s^*)$ from both sides of (5) with $\alpha = \alpha_k$. Next, by letting $k \to \infty$ and using the fact that (e.g. Lemma 3 in LIPPMAN [6])

$$\lim_{k\to\infty} \sup_{a\in A} h_k(a) = \sup_{a\in A} \lim_{k\to\infty} h_k(a)$$

for any sequence $\{h_k(\cdot)\}$ of real-valued functions converging uniformly on A, we get (6) with sup instead of max. This completes the proof. □

For the case where S is compact and A is finite and under assumptions including a special case of A5 the results of Theorem 2 were obtained in YAMADA [15] by using a duality approach. Also under a Doeblin condition on the transition probability functions associated with the stationary policies and some continuity conditions WIJNGAARD [14] proved the existence of an average return optimal policy among the class of stationary policies by using linear perturbation theory. For the case where S is denumerable HORDIJK [4] proved the existence of an average return optimal stationary policy under various assumptions amongst which an assumption of the Doeblin type.

3. EXPONENTIAL CONVERGENCE OF THE VALUE ITERATION METHOD

In this section we shall give conditions under which the value iteration method exhibits exponential convergence. We shall need the result that the optimality equation for the average return criterion has a fixed point. For the case of a finite state and action space SCHWEITZER & FEDERGRUEN [13] proved that the value iteration method exhibits exponential convergence whenever convergence happens. However, according to Markov chain theory, this cannot generally hold when the state space is not finite. For a dynamic inventory model with a continuous state space exponential convergence of the value iteration was established in HORDIJK & TIJMS [5] by exploiting the structure of the model.

Assume now that A1-A2 hold. Let B(S) be the class of all bounded continuous functions on S. Define the mapping T: B(S) → B(S) by

$$(7) \qquad Tu(s) = \max_{a\in A}\{r(s,a) + \int_S u(y)q(dy|s,a)\}, \qquad s \in S.$$

We note that, by the Lemmas 3.4 and 4.1 and the Selection Theorem in MAITRA [8], $Tu \in B(S)$ and there is a $f \in F$ such that $f(s)$ attains the maximum in the right side of (7) for all s. For any $u \in B(S)$, let $\|u\| = \sup_{s\in S} u(s) - \inf_{s\in S} u(s)$.

Given any $u_0 \in B(S)$, define $u_n \in B(S)$ for $n = 1,2,\ldots$ by the value-iteration equation

$$(8) \qquad u_n(s) = \max_{a\in A}\{r(s,a) + \int_S u_{n-1}(y)q(dy|s,a)\}, \qquad s \in S,$$

i.e. $u_n = Tu_{n-1}$. We shall now prove that the sequence $\{u_n(s) - ng\}$ converges exponentially fast and uniformly in s to a function which differs by a constant from the fixed point v(s) of the optimality equation (6). To do this, we introduce the following assumption of the "scrambling" type.

A5'. There are numbers $\rho > 0$, $\gamma > 0$ and for each four elements (s_1,s_2,a_1,a_2) with $s_i \in S$, $a_i \in A$ and $s_1 \neq s_2$ there is a σ-finite measure ϕ on $(S,\mathcal{B})$ and a set $C \in \mathcal{B}$ (ϕ and C may depend on s_i and a_i) such that $\phi(C) \geq \gamma$ and $p(s'|s_i,a_i) \geq \rho$ for all $s' \in C$ and $i = 1,2$ where $p(\cdot|s_i,a_i)$ is the density of the absolutely continuous component of $q(\cdot|s_i,a_i)$ with respect to ϕ.

Although A5' does not imply A5, an examination of the proof given on pp. 197-198 in DOOB [2] shows that, for each $f \in F$, relation (4) also applies under A5'. Since in the analysis of section 2 we used A5 only through (4), the optimality results of section 2 hold also under A5'. We note that A5' includes the case in which there is a number $\rho > 0$ and for each four elements (s_1,s_2,a_1,a_2) with $s_1 \neq s_2$ there is a state s_0 such that $q(\{s_0\}|s_i,a_i) \geq \rho$ for $i = 1,2$ (when S is countable we can say in this case according to Markov chain terminology that for each stationary policy the associated matrix of one-step transition probabilities is scrambling and has an ergodic coefficient of at least ρ).

THEOREM 3. *Suppose that* A1-A4 *and* A5' *hold. Then*

(a) *For all* $u,w \in B(S)$, $\|Tu-Tw\| \leq (1-\rho\gamma)\|u-w\|$, *i.e.* T *is a contraction mapping.*

(b) *For any* $u_0 \in B(S)$, *there are constants* L *and* $\delta = \|u_0-v\|$ *such that*

$$|u_n(s) - ng - v(s) - L| \leq \delta(1-\rho\gamma)^n \qquad \textit{for all } n \geq 1 \textit{ and } s \in S.$$

(c) *For any* $n \geq 1$, *let* $f_n \in F$ *be such that* $f_n(s)$ *attains the maximum in the right side of* (8) *for all* s. *Then, for any* $n \geq 1$,

$$\inf_{s\in S}\{u_n(s)-u_{n-1}(s)\} \leq g(f_n^{(\infty)},s) \leq g \leq \sup_{s\in S}\{u_n(s)-u_{n-1}(s)\}.$$

Moreover, $\sup_s\{u_n(s)-u_{n-1}(s)\}$ *is monotone decreasing to* g *as* $n \to \infty$ *and* $\inf_s\{u_n(s)-u_{n-1}(s)\}$ *is monotone increasing to* g *as* $n \to \infty$, *where the convergence is exponentially fast.*

PROOF. *(a)* Choose $u,w \in B(S)$. Let $f_1 \in F$ and $f_2 \in F$ be such that $f_1(s)$ attains the maximum in the right side of (7) for all s and $f_2(s)$ attains the maximum in the right side of (7) with u replaced by w for all s. Fix $s_0,s_1 \in S$. Now, it easily follows from (7) that

$$(9) \quad (Tu-Tw)(s_0) - (Tu-Tw)(s_1) \leq \int_S \{u(y)-w(y)\}[q(dy|s_0,f_1(s_0))-q(dy|s_1,f_2(s_1))].$$

Next an examination of the proof given on p. 198 in DOOB [2] shows that right side of (9) is less than or equal to $(1-\rho\gamma)\|u-w\|$ which implies part (a) since s_0 and s_1 were arbitrarily chosen.

(b) The proof of part (b) proceeds along standard lines (cf. [5] and [13]). By (6), $T[v + (n-1)\underline{g}] = v + n\underline{g}$ for all $n \geq 1$ where $\underline{g}(\cdot)$ is identical to g on S. A repeated application of this fact and part (a) shows that

$$(10) \qquad \|u_n^*\| \leq (1-\rho\gamma)^n \|u_0-v\| \qquad \text{for all } n \geq 1.$$

Now, let $f \in F$ be as in Theorem 2 and, for $n \geq 1$, let $f_n \in F$ be such that $f_n(s)$ attains the maximum in the right side of (8) for all s. Now, by (6) and (8), we easily get, for all $n \geq 1$,

$$\int_S u_{n-1}^*(y)q(dy|s,f(s)) \leq u_n^*(s) \leq \int_S u_{n-1}^*(y)q(dy|s,f_n(s)) \quad \text{for all } s \in S,$$

so, by induction, we get that $\sup_s u_n^*(s)$ is non-increasing in $n \geq 1$ and $\inf_s u_n^*(s)$ is non-decreasing in $n \geq 1$. From this result and (10) it now follows easily that $\sup_s u_n^*(s)$ and $\inf_s u_n^*(s)$ have a common limit L (say) as $n \to \infty$ and that part (b) holds.

(c) Similarly as in the proof of part (b), we get that $\sup_s\{u_n(s)-u_{n-1}(s)\}$ is non-increasing in $n \geq 1$ and $\inf_s\{u_n(s)-u_{n-1}(s)\}$ is non-decreasing in $n \geq 1$ which proves the second assertion in part (c). By (4), we have for any $f \in F$

$$\pi_f(B) = \int_S q_f^1(B|dy)\pi_f(dy), \quad B \in \mathcal{B}$$

and

$$g(f^{(\infty)},s) = \int_S r(y,f(y))\pi_f(dy), \quad s \in S.$$

Using these relations and making an obvious modification on the argument used on p.243 in HASTINGS [3] we get the other assertion of part (c). □

ACKNOWLEDGEMENT.

I am indebted to Avi Federgruen and Isaac Meilijson for helpful comments.

REFERENCES

[1] BLACKWELL, D., Discounted Dynamic Programming, *Ann. Math. Statist.*, Vol. 36 (1965), 226-235.

[2] DOOB, J.L., *Stochastic Processes*, Wiley, New York, 1953.

[3] HASTINGS, N.A.J., Bounds on the Gain of a Markov Decision Process, *Operations Res.*, Vol. 19 (1971), 240-243.

[4] HORDIJK, A., *Dynamic Programming and Markov Potential Theory*, Mathematical Centre Tract No. 51, Mathematisch Centrum, Amsterdam, 1974.

[5] HORDIJK, A. & H.C. TIJMS, On a Conjecture of Iglehart, *Management Sci.*, Vol. 21 (1975), 1342-1345.

[6] LIPPMAN, S.A., Semi-Markov Decision Processes with Unbounded Rewards, *Management Sci.*, Vol. 19 (1973), 717-731.

[7] LIPPMAN, S.A., On Dynamic Programmning with Unbounded Rewards, *Management Sci.*, Vol. 21 (1975), 1225-1233.

[8] MAITRA, A., Discounted Dynamic Programming on Compact Metric Spaces, *Sankhya*, Vol. 30 (1968), *Ser. A*, 211-216.

[9] ROSS, S.M., Arbitrary State Markovian Decision Processes, *Ann. Math. Statist.*, Vol. 39 (1968), 2118-2122.

[10] ROSS, S.M., *Applied Probability Models with Optimization Applications*, Holden-Day, Inc., San Francisco, 1970.

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

[12] SCHÄL, M., On Continuous Dynamic Programming with Discrete Time-Parameter, *Z. Wahrscheinlichkeitstheorie verw.Gebiete*, Vol. 21 (1972), 279-288.

[13] SCHWEITZER, P.J. & A. FEDERGRUEN, Geometric Convergence of Multichain Value Iteration, 1975 (to appear).

[14] WIJNGAARD, J., *Stationary Markovian Decision Problems, Discrete Time, General State Space*, Thesis, Technology University of Eindhoven, Eindhoven, 1975.

[15] YAMADA, D., Duality Theorem in Markovian Decision Problems, *J. Math. Anal. Appl.*, Vol. 50 (1975), 579-595.