

# Markov programming by successive approximations with respect to weighted supremum norms

**Citation for published version (APA):**

Wessels, J. (1974). *Markov programming by successive approximations with respect to weighted supremum norms*. (Memorandum COSOR; Vol. 7413). Technische Hogeschool Eindhoven.

**Document status and date:**

Published: 01/01/1974

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

7413

ARC  
01  
COS

TECHNOLOGICAL UNIVERSITY EINDHOVEN

Department of Mathematics

STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 74-13

Markov programming by successive approximations  
with respect to weighted supremum norms

by

J. Wessels

Eindhoven, December 1974 (revised)

The Netherlands

Markov programming by successive approximations  
with respect to weighted supremum norms

by

J. Wessels

Summary. Markovian decision processes are considered in the situation of discrete time, countable state space, and general decision space. By introducing a Banach space with a weighted supremum norm, conditions are derived, which guarantee convergence of successive approximations to the value function. These conditions are weaker than those required by the usual supnorm approach. Several properties of the successive approximations are derived.

1. Introduction. We consider a Markov decision process with a countably infinite or finite state space  $\mathcal{S}$  and decision space  $\mathcal{K}$ , defined as follows. A system is observed at discrete points of time ( $t = 0, 1, 2, \dots$ ). If at time  $t$  the state of the system is  $i \in \mathcal{S}$ , a decision  $k \in \mathcal{K}$  may be chosen, which results in a reward  $r_i^k$ . The state  $i$  at time  $t$  and the decision  $k$  determine the probability  $p_{ij}^k$  of observing the system in state  $j$  at time  $t+1$  (regardless of the earlier history of the process). We suppose:

$$\sum_{j \in \mathcal{S}} p_{ij}^k \leq 1 \quad \text{for all } i \in \mathcal{S}, k \in \mathcal{K} .$$

Hence a positive probability for fading of the system is allowed.

A policy  $f$  is a function on  $\mathcal{S}$  with values in  $\mathcal{K}$ . A strategy  $s$  is a sequence of policies:  $s = (f_0, f_1, f_2, \dots)$ . If strategy  $s$  is used, we take decision  $f_t(i)$ , if at time  $t$  the state of the system is  $i$ , i.e. we introduce only so-called (nonrandomized) Markov strategies.

As optimality criterion we choose total expected reward, which is defined for a strategy  $s = (f_0, f_1, \dots)$  as a vector  $V(s)$  in the following way

$$V(s) = \sum_{t=0}^{\infty} [ \prod_{n=0}^{t-1} P(f_n) ] r(f_t) ,$$

where the sum is supposed to remain convergent when rewards are replaced by their absolute values,

$r(f)$  is interpreted as a (column) vector with  $i$ -component  $r_i^{f(i)}$  (for  $i \in \mathcal{S}$ ) for any policy  $f$ , and

$P(f)$  is interpreted as a matrix with  $(i,j)$ -component  $p_{ij}^{f(i)}$  (for  $i,j \in \mathcal{S}$ ) for any policy  $f$ .

Matrix products, matrix-vector products and sums of vectors are defined in the usual way; an empty matrix product is the identical matrix.

This formulation contains the discounted case ( $\beta \leq 1$ ), since the discount-factor may be supposed to be incorporated in the  $p_{ij}^k$ . The same holds for the semi-Markov case, which only requires  $t$  to be interpreted as the number of the decision moment rather than as actual time. For semi-Markov decision processes with discounting the resulting discountfactors depend on  $i,j,k$  and may again be supposed to be incorporated in the  $p_{ij}^k$ .

$V(s)$  converges absolutely and uniformly in its components under the following conditions:

$$\sum_j p_{ij}^k \leq \rho < 1, \quad |r_i^k| \leq M \quad (\text{for all } i \in \mathcal{S}, k \in \mathcal{K}).$$

Under these conditions the total expected reward  $V_i(s)$ , when the system starts in  $i$  and under strategy  $s$ , is at most  $\frac{M}{1-\rho}$  in absolute value. The value function  $V := \sup_s V(s)$  may then be estimated by successive approximations. Upper and lower bounds for  $V$  may be given at each step. At the same time, the method produces at each step a stationary strategy  $s = (f, f, \dots)$  with  $V(s)$  lying between the same bounds. For the finite state, finite decision case this may be found in MacQueen [7], Schellhaas [11], and van Nunen [8]. A more general situation has been treated by Denardo [2].

In this paper we obtain similar results under somewhat weaker conditions, especially the uniformity requirements of the conditions will be weakened. Like Denardo, MacQueen, Schellhaas, and van Nunen, we shall basically apply the contraction operator technique as introduced by Blackwell [1]. However, we shall not use the Banach space of functions on  $\mathcal{S}$  with supremum norm as Blackwell does. We shall introduce a Banach space  $\mathcal{V}$  of functions on  $\mathcal{S}$  with a modified supremum norm. For inventory problems with average costs, Wijngaard [15] introduces a special (exponential) norm of this type. Lippman [6] works

with the same type of norm for the discounted case, however his conditions are more complicated and only guarantee N-stage contraction. Operators in  $\mathbb{W}$  are introduced in section 2. Section 3 presents an approximation procedure for the value function of the problem, together with a procedure to find a strategy which is nearly optimal. In section 4 some possibilities for extensions and for weakening of the conditions are suggested.

2. Norms and operators. Let  $\mu$  be a positive function on  $\mathcal{S}$ , and denote by  $\mathbb{W}$  the set of all real valued functions  $v$  on  $\mathcal{S}$  (interpreted as columnvectors) with the property

$$\|v\| := \sup_{i \in \mathcal{S}} \mu(i) |v(i)| < \infty .$$

As one easily verifies,  $\|\cdot\|$  is a norm and the set  $\mathbb{W}$  is complete with respect to this norm, i.e.  $\mathbb{W}$  is a Banach space.

This norm on  $\mathbb{W}$  induces a norm on the set of real matrices that represent linear operators on  $\mathbb{W}$ , viz.

$$\|A\| := \sup_i \mu(i) \sum_j |a_{ij}| \mu^{-1}(j) .$$

For matrices  $A, B$  with  $\|A\|, \|B\| < \infty$  and  $v \in \mathbb{W}$  we clearly have

$$\|Av\| \leq \|A\| \|v\| \text{ and } \|AB\| \leq \|A\| \|B\| .$$

We now state some assumptions on the reward and probability structure of the system.

Assumptions.

- 1)  $r(f) \in \mathbb{W}$  and  $\|r(f)\| \leq M < \infty$  for all policies  $f$ .
- 2)  $\sup_f \|P(f)\| =: \rho < 1$ .

Assumption 1 means that

$$|r_i^k| \leq \frac{M}{\mu(i)} \quad \text{for all } k \in K \text{ and } i \in \mathcal{S} .$$

Hence, for fixed  $i \in \mathcal{S}$  the rewards for different decisions are bounded, however as a function of  $i$  these bounds may increase to infinity. Actually, a function  $\mu$  exists such that assumption 1 is fulfilled, iff  $r_i^k$  is bounded in  $k$  for fixed  $i$ .

For the probability structure, assumption 2 means, that, given the starting state  $i$  and the decision  $k$ , the expectation of  $\mu^{-1}(X_1)$  is at most  $\rho\mu^{-1}(i)$ , where  $X_1$  is the random variable denoting the state of the system at time  $t = 1$ . In the special case  $\mu \equiv 1$  these assumptions give the well-known conditions mentioned in section 1.

Lemma 1. For any strategy  $s = (f_0, f_1, \dots)$  the total expected reward  $V(s)$  exists, i.e.

$$\sum_{t=0}^{\infty} [\prod_{n=0}^{t-1} P(f_n)] |r(f_t)|$$

converges componentwise and in norm (the vector  $|r|$  has  $i$ -component  $|r_i|$ ),

$$V(s) \in \mathbb{V}, \|V(s)\| \leq \frac{M}{1-\rho} \text{ or } V_i(s) \leq \frac{M}{1-\rho} \mu^{-1}(i) .$$

Proof. The assertion follows from the fact that

$$v_t := [\prod_{n=0}^{t-1} P(f_n)] |r(f_t)| \in \mathbb{V} ,$$

with  $\|v_t\| \leq \rho^t M$ . □

On  $\mathbb{V}$  we define the operators  $L_f$  for any policy  $f$  and  $U$ , by

$$L_f v := r(f) + P(f)v \quad \text{for any } v \in \mathbb{V}$$

$$(Uv)(i) := \sup_{k \in K} \{r_i^k + \sum_j p_{ij}^k v(j)\} \quad \text{for } i \in S, v \in \mathbb{V} ,$$

or in matrix notation:

$$Uv := \sup_f \{r(f) + P(f)v\} ,$$

where the sup is taken componentwise.

Lemma 2.

- a)  $L_f$  and  $U$  map  $\mathbb{V}$  into  $\mathbb{V}$ .
- b)  $L_f$  and  $U$  are monotone mappings.
- c)  $L_f$  and  $U$  map  $\{v \in \mathbb{V} \mid \|v\| \leq \frac{M}{1-\rho}\}$  into itself.
- d)  $L_f$  and  $U$  are strictly contracting with contraction radii  $\|P(f)\|$  and  $\rho$  respectively.

- e)  $L_f$  and  $U$  possess unique fixed points in  $\mathbb{V}$  with norms at most  $\frac{M}{1-\rho}$ .  
 f) the fixed point of  $L_f$  is  $V(f^{(\infty)})$ , where  $f^{(\infty)}$  denotes the stationary strategy  $(f, f, f, \dots)$ .

Proof. The proofs of a), b), c) are straightforward. For the finite state, finite decision case with  $\mu \equiv 1$  property c) has been noticed by Shapiro [12]. e) is a direct consequence of d) and assumption 2. f) is proved by direct verification. About d) the following remarks. The proof of the fact that  $L_f$  is strictly contracting with contraction radius at most  $\|P(f)\|$  is straightforward. The example  $v(i) := \mu^{-1}(i)$ ,  $w(i) := 0$  ( $i \in \mathcal{S}$ ) shows that for certain  $v, w \in \mathbb{V}$

$$\|L_f v - L_f w\| = \|P(f)\| \|v - w\|.$$

That  $U$  has contraction radius at most  $\rho$  is proved in the following way. Choose  $v \in \mathbb{V}$  and  $\varepsilon > 0$ . For any  $i \in \mathcal{S}$  a decision  $k$  is chosen such that

$$r_i^k + \sum_j p_{ij}^k v(j) \geq (Uv)(i) - \varepsilon \mu^{-1}(i).$$

Now for this  $v$  and an arbitrary  $w \in \mathbb{V}$  we have

$$\begin{aligned} \mu(i)(Uv)(i) - \mu(i)(Uw)(i) &\leq \mu(i)r_i^k + \mu(i) \sum_j p_{ij}^k v_j + \varepsilon + \\ &- \mu(i)r_i^k - \mu(i) \sum_j p_{ij}^k w(j) = \varepsilon + \mu(i) \sum_j p_{ij}^k (v(j) - w(j)) \leq \\ &\leq \varepsilon + \rho \|v - w\|. \end{aligned}$$

In the same way we prove for arbitrary  $v$  and  $w$

$$\mu(i)(Uw)(i) - \mu(i)(Uv)(i) \leq \varepsilon + \rho \|v - w\|.$$

Hence

$$\|Uv - Uw\| \leq \varepsilon + \rho \|v - w\| \quad \text{for all } \varepsilon > 0,$$

and therefore

$$\|Uv - Uw\| \leq \rho \|v - w\|.$$

By substituting  $w(i) := 0$ ,  $v(j) := \ell \mu^{-1}(j)$  with  $\ell > 0$ , we verify that

$$\|Uv - Uw\| \geq [-2\varepsilon + \rho] \|v - w\| \quad \text{if } \ell \geq \frac{M}{\varepsilon}.$$

This implies that  $\rho$  is the contraction radius. □

3. Approximation procedures. Fixed points of contraction mappings on  $\mathbb{W}$  may be approximated by a sequence of points in  $\mathbb{W}$ . For the operator  $U$ , such a sequence is generated in the following way: choose  $v_0 \in \mathbb{W}$ , define recursively  $v_n := Uv_{n-1}$  for  $n = 1, 2, \dots$ . Then  $v_n$  converges in norm to the fixed point  $w$  of  $U$ :  $\lim_{n \rightarrow \infty} \|v_n - w\| = 0$ , or, for  $\epsilon > 0$  there exists a number  $N_\epsilon$  such that for  $n \geq N_\epsilon$ :

$$|v_n(i) - w(i)| < \epsilon \mu^{-1}(i), \quad \text{for all } i \in \mathbb{S}.$$

As  $U$  is monotone, we obtain a nondecreasing sequence, if  $v_0$  is chosen such that  $v_0 \leq v_1$ .

This can be achieved by taking  $v_0 := -\frac{M}{1-\rho} \mu^{-1}$ , where  $\mu^{-1} \in \mathbb{V}$  with components  $\mu^{-1}(i)$ . By assumptions 1 and 2 we then have

$$\begin{aligned} v_1 &= \sup_f \{r(f) - P(f) \frac{M}{1-\rho} \mu^{-1}\} \geq -M\mu^{-1} - \frac{M}{1-\rho} \rho \mu^{-1} = \\ &= -\frac{M}{1-\rho} \mu^{-1} = v_0. \end{aligned}$$

It seems natural to conjecture that  $w = V (= \sup_s V(s))$ . We first prove

Lemma 3. For any strategy  $s = (f_0, f_1, \dots)$  we have  $V(s) \leq w$ , i.e.  $V \leq w$ .

Proof.

$$V(s) = \sum_{t=0}^{\infty} \left[ \prod_{n=0}^{t-1} P(f_n) \right] r(f_t) \leq \sum_{t=0}^{N-1} \left[ \prod_{n=0}^{t-1} P(f_n) \right] r(f_t) + \sum_{t=N}^{\infty} \rho^t M \mu^{-1}.$$

Hence if  $\frac{\rho^N M}{1-\rho} \leq \epsilon$ , i.e. if  $N$  sufficiently large, we have

$$V(s) \leq U^N 0 + \epsilon \mu^{-1},$$

where  $0$  denotes the element of  $\mathbb{W}$  with all components  $0$ .

$U^N 0$  converges in norm and hence componentwise to  $w$  when  $N \rightarrow \infty$ . This implies

$$V(s) \leq w + \epsilon \mu^{-1}.$$

This inequality holds componentwise for any  $\epsilon > 0$ , hence  $V(s) \leq w$ . □



Theorem 1. For any  $\epsilon > 0$  there is a policy  $f$ , such that the stationary strategy  $f^{(\infty)} := (f, f, \dots)$  satisfies

$$\|V(f^{(\infty)}) - w\| \leq \epsilon ,$$

hence  $V = w$ .

Proof. Let  $\delta := \frac{1}{2}(1 - \rho)\epsilon$ . Select  $v_0 \in W$ , such that  $v_0 < Uv_0$  (strictly smaller for each component, e.g.  $v_0 = -c\mu^{-1}$  with  $c < \frac{M}{1 - \rho}$ ). A policy  $f_n$  ( $n = 1, 2, \dots$ ) is selected, such that

$$v_n := L_{f_n} v_{n-1} \geq \max\{v_{n-1}, Uv_{n-1} - \delta\mu^{-1}\} ,$$

where the maximum is taken componentwise. Such a policy  $f_n$  can always be found, as can be seen as follows. If  $v_{n-1}(i) < (Uv_{n-1})(i)$  it is trivial by the definition of  $U$ . If  $v_{n-1}(i) = (Uv_{n-1})(i)$  for certain  $i \in S$ , then  $f_n(i) = f_{n-1}(i)$  satisfies, because - using induction - we have

$$v_{n-1} = L_{f_{n-1}} v_{n-2} \geq v_{n-2} ,$$

hence

$$(L_{f_{n-1}})^2 v_{n-2} \geq v_{n-1} , \text{ or } L_{f_{n-1}} v_{n-1} \geq v_{n-1} ,$$

as required.

We now proceed with the proof. The same reasoning gives

$$(L_{f_n})^k v_{n-1} \geq v_n ,$$

for any natural number  $k$ . Hence

$$V \geq V(f_n^{(\infty)}) \geq v_n .$$

It now suffices to prove that  $v_n$  approximates  $w$  in norm for sufficiently large  $n$ . We have

$$\begin{aligned} v_n = L_{f_n} v_{n-1} &\geq Uv_{n-1} - \delta\mu^{-1} \geq U[Uv_{n-2} - \delta\mu^{-1}] - \delta\mu^{-1} \geq \\ &\geq U^2 v_{n-2} - \delta\rho\mu^{-1} - \delta\mu^{-1} . \end{aligned}$$

Repetition of this argument yields

$$v_n \geq U^n v_0 - \delta(1 + \rho + \dots + \rho^{n-1})\mu^{-1} \geq U^n v_0 - \frac{\delta}{1 - \rho} \mu^{-1} .$$

Summarizing we have

$$U^n v_0 - \frac{\delta}{1-\rho} \mu^{-1} \leq v_n \leq V(f_n^{(\infty)}) \leq V \leq w .$$

Since  $U^n v_0$  converges to  $w$  (in norm), we have for  $n$  sufficiently large

$$\|w - V(f_n^{(\infty)})\| \leq \frac{2\delta}{1-\rho} = \varepsilon . \quad \square$$

Now we have proved that the fixed point  $w$  of the operator  $U$  is equal to the optimal value vector  $V$  of the decision problem. Furthermore we have proved that for any  $\varepsilon > 0$  a stationary strategy is  $\varepsilon$ -optimal (defined in terms of the norm). The question now arises whether one is able to find lower and upper bounds for  $V(f_n^{(\infty)})$  and  $V$  at the  $n$ -th iteration step of the iteration process developed in the proof of theorem 1. Apparently,  $v_n$  is a lower bound. However, without much effort a better lower bound and upper bound can be constructed. The proofs follow the same line as van Nunen's proof [8] for the bounds of MacQueen [7] in the  $\mu \equiv 1$ , finite state, finite decision case. The same technique turned out to work for a variety of other successive approximation methods for the same case (van Nunen [8], van Nunen [9], Wessels and van Nunen [14]). Hinderer [4] used a similar approach for finite horizon problems.

Theorem 2. Suppose  $\delta > 0$ ;  $v, w \in \mathbb{V}$  such that  $Uw - \delta\mu^{-1} \leq v$ . Then

$$V \leq v + \frac{\delta + \rho\|v - w\|}{1-\rho} \mu^{-1} .$$

Proof.  $Uv = U(w + v - w)$ . Hence, since  $Uw \leq v + \delta\mu^{-1}$

$$Uv \leq Uw + \rho\|v - w\|\mu^{-1} \leq v + \delta\mu^{-1} + \rho\|v - w\|\mu^{-1} .$$

This implies  $Uv \leq v + \varepsilon\mu^{-1}$ , with  $\varepsilon := \delta + \rho\|v - w\|$ . Hence

$$\begin{aligned} U^2 v &\leq U(v + \varepsilon\mu^{-1}) = U(w + v - w + \varepsilon\mu^{-1}) \\ &\leq Uw + \rho\|v - w\|\mu^{-1} + \varepsilon\rho\mu^{-1} \\ &\leq v + \delta\mu^{-1} + \rho\|v - w\|\mu^{-1} + \varepsilon\rho\mu^{-1} . \end{aligned}$$

Or

$$U^2 v \leq v + \varepsilon(1 + \rho)\mu^{-1} .$$

Generally

$$U^N v \leq v + \frac{\varepsilon}{1 - \rho} \mu^{-1} ,$$

which implies, since  $\lim_{N \rightarrow \infty} U^N v = V$ :

$$V \leq v + \frac{\varepsilon}{1 - \rho} \mu^{-1} .$$

□

Theorem 3. If  $v, w \in \mathbb{V}$  satisfy  $L_f w \geq v$ , then

$$v + \frac{\rho_* \|v - w\|_*}{1 - \rho_*} \mu^{-1} \leq V(f^{(\infty)}) \leq v + \frac{\rho \|v - w\|}{1 - \rho} \mu^{-1} ,$$

where

$$\|v - w\|_* := \inf_i \mu(i)(v(i) - w(i)) ,$$

$$\rho_* := \inf_{i,k} \mu(i) \sum_j p_{ij}^k \mu^{-1}(j) .$$

The proof proceeds as the proof of theorem 2.

□

Remark. In theorem 3 the values of  $\rho$  and  $\rho_*$  may be replaced by

$$\rho(f) := \|P(f)\|$$

and

$$\rho_*(f) := \inf_i \mu(i) \sum_j p_{ij}^{f(i)} \mu^{-1}(j)$$

respectively. These replacements make the assertions sharper, however, they take more work.

We have now proved, that the following algorithm ends after a finite number of steps:

start: choose  $\alpha > 0$ ,  $\delta > 0$ ,  $v_0 \in \mathbb{V}$  with  $v_0 < Uv_0$  (< for all components) and  $\frac{\delta}{1 - \rho} < \alpha$ .

iteration part: find for  $n = 1, 2, \dots$  a policy  $f_n$ , such that

$$v_n := L_{f_n} v_{n-1} \geq \max\{v_{n-1}, Uv_{n-1} - \delta\mu^{-1}\} ,$$

until

$$\frac{\delta + \rho \|v_n - v_{n-1}\|}{1 - \rho} - \frac{\rho_* \|v_n - v_{n-1}\|_*}{1 - \rho_*} < \alpha .$$

stop:

$$v_n + \frac{\rho_* \|v_n - v_{n-1}\|_*}{1 - \rho_*} \mu^{-1} \leq V(f_n^{(\infty)}) \leq V \leq v_n + \frac{\delta + \rho \|v_n - v_{n-1}\|}{1 - \rho} \mu^{-1} .$$

with a distance between lower and upper bound of less than  $\alpha$ .

$$V(f_n^{(\infty)}) \leq v_n + \frac{\rho \|v_n - v_{n-1}\|}{1 - \rho} \mu^{-1} ,$$

hence the distance between upper and lower bounds for

$$V(f_n^{(\infty)}) \text{ is less than } \alpha - \frac{\delta}{1 - \rho} .$$

4. Extensions and remarks. An interesting extension of the theory presented here, these spaces and norms could be used to develop analogues to other successive approximation methods. For the supnorm case different successive approximation methods have been proposed (e.g. Reetz [10], Schellhaas [11], van Nunen [8]). These and several other ideas have been combined and extended by van Nunen [9], whereas a more general approach for generating successive approximation procedures for the supnorm case has been presented by Wessels [13] and Wessels and van Nunen [14]. In the papers [8] and [14], Howard's policy iteration method [5] appears as a specific successive approximation procedure. It seems possible to weaken the conditions under which these methods work.

An other interesting situation for extension in the sense of this paper may be found in a paper by Harrison [3]. Harrison considers a situation with unbounded reward functions where successive approximations converge in supremum norm if the starting vector is well chosen.

In the present paper the condition is:

A: a positive function  $\mu$  exists, such that assumptions 1 and 2 are satisfied.

For the finite state case ( $\mathcal{S}$  finite) condition A is equivalent to

B:  $r_i^k$  is bounded as a function of  $i, k$ , and there exist a positive number  $\epsilon$  and a natural number  $N$ , such that

$$P(X_N \in \mathcal{S} \mid X_0 = i, \text{ strategy } s) \leq 1 - \epsilon \quad \text{for all } s, i .$$

The proof of the equivalence is rather straightforward.

Actually, B implies A if  $\mathcal{S}$  is countably infinite, which is proved in the same way as in the finite case.

Such topics will be treated more extensively in a forthcoming paper by K.M. van Hee and the present author.

Condition A may be weakened by replacing assumption 2 by 2\*.

Assumption 2\*. For some  $T \geq 1$ , and all  $f_0, f_1, \dots, f_{T-1}$

$$\| \prod_{t=0}^{T-1} P(f_t) \| \leq \rho < 1 .$$

It is not necessary to use a fixed  $\delta$  in the algorithm: the  $\delta$ -value,  $\delta_n$  say, used in the  $n$ -th situation, may depend on  $n$ ; it is only required that  $\delta_n \leq \delta^* < \alpha(1 - \rho)$ , for  $n$  sufficiently large.

### References

- [1] D. Blackwell, Discounted dynamic programming. *Ann. Math. Statist.* 36 (1965), 226-234.
- [2] E.V. Denardo, Contraction mappings in the theory underlying dynamic programming. *SIAM Review* 9 (1967), 165-177.
- [3] J.M. Harrison, Discrete dynamic programming with unbounded rewards. *Ann. Math. Statist.* 43 (1972), 636-644.
- [4] K. Hinderer, Estimates for finite-stage dynamic programs. To appear in *J. Math. Anal. Appl.*
- [5] R.A. Howard, *Dynamic programming and Markov processes*. MIT-Press, Cambridge, 1960.
- [6] S.A. Lippman, On dynamic programming with unbounded rewards. *Management Science* 21 (1975), 1225-1233.
- [7] J. MacQueen, A modified dynamic programming method for Markovian decision problems. *J. Math. Anal. Appl.* 14 (1966), 38-43.
- [8] J.A.E.E. van Nunen, A set of successive approximation methods for discounted Markovian decision problems. To appear in *Z. f. Oper. Res.* 19 (1975).

- [9] J.A.E.E. van Nunen, Improved successive approximation methods for discounted Markov decision processes. To appear in Colloquia Mathematica Societatis János Bolyai 12 (A. Prékopa ed.) North-Holland publ. co. Amsterdam.
- [10] D. Reetz, Solution of a Markovian decision problem by successive over-relaxation. Z. f. Oper. Res. 17 (1973), 29-32.
- [11] H. Schellhaas, Zur Extrapolation in Markoffschen Entscheidungsmodellen mit Diskontierung. Z. f. Oper. Res. 18 (1974), 91-104.
- [12] J.F. Shapiro, Brouwer's Fixed Point Theorem and Finite State Space Markovian Decision Theory. J. Math. Anal. Appl. 49 (1975), 710-712.
- [13] J. Wessels, Stopping times and Markov programming. To appear in Proceedings of the 1974 European Meeting of Statisticians and 7th Prague Conference on Information Theory, Statistical Decision functions, and Random Processes.
- [14] J. Wessels and J.A.E.E. van Nunen, A principle for generating optimization procedures for discounted Markov decision processes. To appear in Colloquia Mathematica Societatis János Bolyai 12 (A. Prékopa ed.) North-Holland publ. co Amsterdam.
- [15] J. Wijngaard, Stationary Markovian decision problems; discrete time, general state space. Eindhoven 1975.