Markov decision processes and strongly excessive functions

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics

PROBABILITY THEORY, STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 77-11

Markov decision processes and strongly excessive functions

by

K.M. van Hee and J. Wessels

Key words: Markov decision process, excessive function, transient behaviour, exponentially bounded stopping time, spectral radius.

This paper is a revised and extended version

of Memorandum COSOR 75-22

Eindhoven, May 1977

The Netherlands

Markov decision processes and strongly excessive functions

by

K.M. van Hee and J. Wessels

0. Summary.

Strongly excessive functions play an important role in the theory of Markov decision processes and Markov games. In this paper the following question is investigated: What are the probabilistic properties of Markov decision processes which posses a strongly excessive function? A characterization is presented in the form of a random drift through a partitioned state space. For strongly excessive functions which have a positive lower bound a characterization is given in terms of the lifetime distribution of the process. Finally we give a characterization in terms of the spectral radius.

1. Introduction.

When analyzing (semi-) Markov decision processes and Markov games one often applies contraction properties of certain operators in a Banach space. This technique has been introduced by Blackwell [1], using 1) the boundedness of the immediate return in the supremum norm and 2) discounting, which is equivalent to a positive probability β of leaving the system in each state (for all strategies). The idea has been generalized by Denardo [2] who weakened the discounting condition by assuming a positive probability of leaving the system in N stages (uniform in the starting state and the strategy). In order to obtain weaker conditions other norms might be used. Norms which appear to be useful are of the weighted supremum norm type. First attempts in this direction have been made by Veinott [10] in case of a finite state space and by Lippman [7] using a polynomial and a special condition on the transition probabilities. More general approaches have been presented by Lippman [8] and by one of the present authors [13]. Hinderer [5] uses a similar technique as [13] for finite stage programs. Wijngaard [14, ch. 5] uses weighted supremum norms (with an exponential weight function) for analyzing average costs inventory problems.

In this paper we will investigate the probabilistic properties of the decision processes when the transition probabilities satisfy the conditions imposed by the weighted supremum norm approach. These conditions may be formulated (see below) as the existence of a function on the state space which is excessive in a somewhat stronger sense than usual (compare Hordijk [6]).

In section 3 we give a characterization of the existence of a strongly excessive function in the form of a random drift through a partition of the state space and in section 4 these properties are related to the lifetime distribution of the process.
Further we give in section 5, an analytic equivalent for the existence of a strongly excessive function in terms of the spectral radius of the decision process.

A *Markov decision process* is determined by a pair $(E,\mathcal{P})$ where E is called the *state space* (supposed to be countable in this paper), $\mathcal{P}$ is a set of sub-Markov matrices ($P \in \mathcal{P}$ is a nonnegative function on $E \times E$ with $\sum_{j \in E} P(i,j) \le 1$ for all $i \in E$). It is usual to define a Markov decision process as a triple $(E,\mathcal{P},r)$ where r is a real function on $E \times \mathcal{P}$ with the interpretation of a reward function. However in this paper we are only dealing with the state space E and the transition probabilities $\mathcal{P}$.
At this moment we do not require any structure on $\mathcal{P}$ but in section 2 we make an assumption which is filled if we are dealing with the usual law of motion of a Markov decision process (see e.g. Blackwell [1]). Consider a positive function μ on E and introduce the Banach space V_μ of all real valued functions v on E which satisfy

$$\| v \|_\mu := \sup_{i \in E} \frac{v(i)}{\mu(i)} < \infty .$$

We call $\| \cdot \|_\mu$ the *weighted supremum norm* and μ a *bounding function*. The norm concept in V_μ induces a norm for the matrices $P \in \mathcal{P}$, viz. the operator norm

$$\| P \|_\mu := \sup_{i \in E} \mu(i)^{-1} \sum_{j \in E} P(i,j)\mu(j) .$$

Definition 1.

A Markov decision process $(E,\mathcal{P})$ is called *contracting* if there exists a bounding function μ on E and a number $0 < \rho < 1$ such that

$$\| P \|_\mu \le \rho \quad \text{for all } P \in \mathcal{P} .$$

In [13] the *contracting dynamic programming* model is analysed extensively.
Note that $\| P \|_\mu \le \rho < 1$ for all $P \in \mathcal{P}$ is equivalent to

$$P\mu(i) := \sum_{j \in E} P(i,j)\mu(j) \le \rho\mu(i) \qquad \text{for all } P \in \mathcal{P} .$$

For $\rho = 1$ this condition becomes the usual requirement for excessivity of the function μ (see Hordijk [6]). So our condition is stronger.

Definition 2.

A bounding function v on E is called *strongly excessive* with respect to $(E,\mathcal{P})$ if there is a number ρ $(0 < \rho < 1)$ with $Pv \le \rho v$ for all $P \in \mathcal{P}$.
The number ρ is called an *excessivity factor*.

Remark.

The contracting condition may be used in the total expected rewards case and in the total expected discounted reward case, viz. if in the discounted case Q is a transition matrix we define a matrix P as βQ where β is the discountfactor $(0 \le \beta < 1)$.
In the same way discounted semi-Markov decision processes may be handled by defining

$$\beta_Q(i,j) := \int_{0-}^{\infty} e^{-\alpha t} dF_Q(t;i,j) \ , \quad P(i,j) := \beta_Q(i,j)Q(i,j) \ .$$

2. Prerequisites and notations.

A *Markov strategy* R is a sequence $(P_0,P_1,\dots)$ of elements of $\mathcal{P}$. The set of all Markov strategies is denoted by M.
We extend the statespace E to $\bar{E}$ by adding a new state x in the following way: $P(x,x) := 1$, $P(i,x) := 1 - \sum_{j\in E} P(i,j)$ for all $i \in E$, for all $P \in \mathcal{P}$.
All functions on E are extended to functions on $\bar{E}$ by defining them 0 in x.
(Note that a strongly excessive function v on $(E,\mathcal{P})$ with excessivity factor ρ satisfies $\rho v \ge Pv$ on $\bar{E}$ for all $P \in \mathcal{P}$).
Any starting $i \in \bar{E}$ and any $R \in M$ determine a (nonhomogeneous)-Markov chain on $\bar{E}$ and so probability $\mathbb{P}_{i,R}$ on $(\bar{E})^{\infty}$. Let $\mathbb{E}_{i,R}$ be the expectation with respect to $\mathbb{P}_{i,R}$. The functions X_n defined by $X_n(\omega) := \omega(n)$ for $\omega = (\omega(1),\omega(2),\dots) \in (\bar{E})^{\infty}$ form a Markov chain. The following lemmas will be used in section 3.

Lemma 1.

Suppose $A \subset E$, $i \in E$, $R_0 \in M$. Then

$$\sum_{n=0}^{\infty} \mathbb{P}_{i,R_0}[X_n \in A] \le \sup_R \sup_{j\in A} \sum_{m=0}^{\infty} \mathbb{P}_{j,R}[X_m \in A] \ .$$

Proof.

The assertion is trivial for $i \in A$. Suppose $i \notin A$. Using the Markov property we obtain

$$\sum_{n=0}^{\infty} \mathbb{P}_{i,R_0}[X_n \in A] \le \sup_R \sum_{n=0}^{\infty} \sum_{k=1}^{n} \sum_{j\in A} \mathbb{P}_{i,R_0}[X_k = j, X_\ell \notin A \quad (1 \le \ell < k)] \; \mathbb{P}_{j,R}[X_{n-k} \in A]$$

$$\le \sup_R \sum_{j\in A} \left(\sum_{m=0}^{\infty} \mathbb{P}_{j,R}[X_n \in A] \right) \left(\sum_{k=1}^{\infty} \mathbb{P}_{i,R_0}[X_k = j, X_\ell \notin A \; (1 \le \ell < k)] \right) \le \sup_R \sup_{j\in A} \sum_{m=0} \mathbb{P}_{j,R}[X_m \in A].$$

Lemma 2.

Suppose (E,P) is contracting with excessivity factor $\rho < 1$. Suppose $1 < \alpha < \rho^{-1}$. Then there exists a strongly excessive function c for (E,P) with excessivity factor $\alpha\rho$, such that c maps E into the set of $\{\alpha^\ell \mid \ell \in Z\}$. *)

Proof.

Define $c(i) := \alpha^\ell$ if $\alpha^{\ell-1} < b(i) \le \alpha^\ell$, where b is a strongly excessive function for (E,P) with excessivity factor ρ. Hence $b \le c$.
If $c(i) = \alpha^\ell$, then $c(i) - b(i) \le \alpha^\ell - \alpha^{\ell-1} = c(i)\,(1 - \alpha^{-1})$. Hence
$Pc = Pb + P(c - b) \le \rho b + (1 - \alpha^{-1})Pc \le \rho c + (1 - \alpha^{-1})Pc$. This implies
$\alpha^{-1}Pc \le \rho c$. □

Lemma 2 allows us to consider only exponentials as candidates for strongly excessive functions. If there is a strongly excessive function that is bounded away from zero the same holds for equidistant grids:

Lemma 3.

Suppose b is a strongly excessive function for (E,P) satisfying $0 < \delta \le b(i)$ for $i \in E$. Suppose $\rho + \alpha\delta^{-1} < 1$, where $\alpha > 0$ and ρ is an excessivity factor for b. Then there exists a strongly excessive function c for (E,P) with excessivity factor $\rho + \alpha\delta^{-1}$, such that c maps E into the set $\{\delta + \ell\alpha \mid \ell = 1,2,\ldots\}$.

Proof.

Define $c(i) = \delta + \ell\alpha$ if $\delta + (\ell - 1)\alpha \le b(i) < \delta + \ell\alpha$.
Hence $c(i) - \alpha \le b(i) < c(i)$, which implies

*) Z is the set of integers.

$$\rho c > \rho b \geq Pb \geq Pc - \alpha Pe \geq Pc - \alpha e \geq Pc - \alpha\delta^{-1}c \ ,$$

where $e(i) = 1$ for all $i \in E$. Therefore $Pc \leq (\rho + \alpha\delta^{-1})c$. □

We now introduce an assumption for $\mathcal{P}$ which will be supposed to hold throughout the rest of this paper.

Assumption.

Let $P_1, P_2, P_3, \ldots \in \mathcal{P}$ and let $A_1, A_2, A_3, \ldots$ be a partition of E. It holds that P defined by

$$P(i,\cdot) := P_j(i,\cdot) \ , \qquad \text{if } i \in A_j$$

is also an element of .

In section 2,3 we need a special form of Bellman's optimality equation:

Lemma 4.

Suppose r is a nonnegative function on E, λ is positive, and

$$v(i) := \sup_{R \in M} \sum_{n=0}^{\infty} \lambda^n \mathbb{E}_{i,R}[r(X_n)] < \infty \ .$$

Then v satisfies for any $P \in \mathcal{P}$ $v \geq r + \lambda P v$.

Note that $\lambda > 1$ is allowed. The proof is a direct consequence of a theorem proved in [4]. Here the assumption is used.

Lemma 5.

Consider a Markov decision process $(E,\mathcal{P})$.

If there is a function $w \geq 0$ on E such that one of the following conditions hold for all $i \in E$

(i) $v(i) := \sup_R \mathbb{E}_{i,R}[w(X_n)] < \infty$ and $w(i) \geq (1 - \rho)\, v(i)$ $\quad (0 < \rho < 1)$,

(ii) $z(i) := \sup_R \mathbb{E}_{i,R}[\beta^n w(X_n)] < \infty$, $\quad \beta > 1$,

then there is strongly excessive function for $(E,\mathcal{P})$.

Proof:

Suppose i). By lemma 4 (with $\lambda = 1$) we have $v \geq w + Pv$ hence $v \geq (1 - \rho)v + Pv$ and therefore $\rho v \geq Pv$.

Suppose ii). Again by lemma 4 (with $\beta = \lambda$) we have $z \geq w + \beta P z$ hence $\frac{1}{\beta} z \geq Pz$. □

In the next sections we shall search for functions $w \geq 0$ on E satisfying one of the conditions i) and ii) of lemma 5 to prove the existence of strongly excessive functions.

3. Probabilistic equivalent for strong excessivity.

In this section it is shown that strong excessivity is related to certain drifting properties of the Markov chain involved.

Theorem 1.

$(E,\mathcal{P})$ is contracting, if and only if there exist a partition $\{E_k | k \in Z\}$ of E and numbers $\alpha > 1$, $\beta \geq 1$, such that for all $R \in M$

$$\sum_{n=0}^{\infty} \mathbb{P}_{i,R}[X_n \in E_k] \leq \beta \min\{1,\alpha^{\ell-k}\} \qquad \text{if } i \in E_\ell \ .$$

Proof:

We first prove the "if"-part, hence we suppose a partition of E with the properties mentioned in the assertion. Choose ε with $0 < \varepsilon < 1$, $\varepsilon\alpha > 1$. The positive functions r and v on E are defined by

$$r(i) := (\varepsilon\alpha)^k \qquad \text{if } i \in E_k \ , \ v(i) := \sup_R \sum_{n=0}^{\infty} \mathbb{E}_{i,R}[r(X_n)] \ .$$

Then for $i \in E_\ell$

$$v(i) = \sup_R \sum_{k \in Z} (\varepsilon\alpha)^k \sum_{n=0}^{\infty} \mathbb{P}_{i,R}[X_n \in E_k]$$

$$\leq \beta \sum_{k<\ell} (\varepsilon\alpha)^k + \beta \sum_{k\geq\ell} (\varepsilon\alpha)^k \alpha^{\ell-k} = \beta(\varepsilon\alpha)^\ell \{(\varepsilon\alpha - 1)^{-1} + (1-\varepsilon)^{-1}\}$$

$$= \beta r(i)\{(1-\varepsilon)^{-1} + (\varepsilon\alpha - 1)^{-1}\}$$

$$= (1-\rho)^{-1} r(i), \quad \text{for certain } \rho \text{ with } 0 < \rho < 1.$$

Hence, according to lemma 5, we have for all $P \in \mathcal{P}$ $\rho v \geq Pv$.

Now the "only if"-part will be proven, hence it is assumed that $(E,\mathcal{P})$ is contracting with excessivity factor $\rho < 1$. Without loss of generality, we may assume that the strongly excessive function b is equal to α^ℓ for $i \in E_\ell$, where $\{E_\ell \mid \ell \in Z\}$ is some partition on E and $\alpha > 1$ (lemma 2)$\cdot \rho b \geq Pb$ for all $P \in \mathcal{P}$ implies for any Markov strategy $R = (P_0, P_1, \ldots)$

$$\rho^n b \geq P_0 \ldots P_{n-1} b \ , \quad \text{or} \quad (1-\rho)^{-1} b \geq \sum_{n=0}^{\infty} \left(\prod_{m=0}^{n-1} P_m \right) b \ ,$$

(where an empty product is equal to the unit matrix).
For $i \in E_\ell$ this means

$$(1-\rho)^{-1} \alpha^\ell \geq \sum_{k \in Z} \alpha^k \sum_{n=0}^{\infty} \mathbb{P}_{i,R}[X_n \in E_k] \ , \text{ hence}$$

$$(*) \qquad (1-\rho)^{-1} \alpha^\ell \geq \alpha^k \sum_{n=0}^{\infty} \mathbb{P}_{i,R}[X_n \in E_k] \ .$$

With $\beta := (1-\rho)^{-1}$ this settles the assertion for $k \geq \ell$.
For $k < \ell$ we apply lemma 1 with $A = E_k$:

$$\sum_{n=0}^{\infty} \mathbb{P}_{i,R}[X_n \in E_k] \leq \sup_{R_0} \sup_{j \in E_k} \sum_{m=0}^{\infty} \mathbb{P}_{j,R_0}[X_m \in E_k] \leq (1-\rho)^{-1} \ .$$

where the inequality (*) has been used with $\ell = k$. □

Remark.

In the construction of a strongly excessive function in the first part of the proof, the choice of ε determines the excessivity factor ρ. For $\varepsilon = \alpha^{-\frac{1}{2}}$ the constructed ρ is minimal, viz. $1 - (\alpha^{\frac{1}{2}} - 1)\beta^{-1}(\alpha + 1)^{-1}$.

4. Strong excessivity and exponentially bounded life times.

Since we supposed $\sum_j P(i,j) \leq 1$, there may be positive probabilities for certain i that the process does not exist any more after one step. Hence we can speak of the life time T of the process. We say that $T \geq n$ iff $X_n \in E$, hence for $R = (P_0, P_1, \ldots)$

$$\mathbb{P}_{i,R}[X_n \in E] = \mathbb{P}_{i,R}[T \geq n] = (P_0 P_1 \ldots P_{n-1} e)(i)$$

where is the unit function on E.
In this section we will investigate the relation between strong excessivity and exponential boundedness of the life time distributions.

Definition 2.

1. The life time T of (E,P) is said to be *exponentially bounded* iff there exist a real number $\gamma (0 < \gamma < 1)$ and a positive function a on E with for all $R \in M$ and $i \in E$

$$\mathbb{P}_{i,R}[T \geq n] \leq a(i)\gamma^n$$

2. If the function a in definition 2.1 does not depend on i, the life time T of (E,P) is said to be *uniformly exponentially bounded*.

Remarks.

1. In the case of discounting we have $P(i,j) := \beta Q(i,j)$ with $0 < \beta < 1$, $\sum_j Q(i,j) \leq 1$. Hence the life time is uniformly exponentially bounded with $a(i) = 1$, $\gamma = 1 - \beta$.

2. In the case of discounted semi-Markov decision processes we have

$$\beta_Q(i,j) := \int_{0-}^{\infty} e^{-\alpha t} dF_Q(t;i,j), \; P(i,j) := \beta_Q(i,j)Q(i,j), \; \sum_j Q(i,j) \leq 1.$$

Hence the life time is uniformly exponentially bounded if

$$\beta_Q(i,j) \leq \beta < 1 \; .$$

3. In the nondiscounted case we have: if there exist a natural number M and a real number $\varepsilon > 0$, such that for all R,i

$$\mathbb{P}_{i,R}[X_M \in E] \leq 1 - \varepsilon,$$

then (E,P) is uniformly exponentially bounded.

For the next lemma we do not need our assumption for P. This lemma seems to be well-known in statistical sequential analysis (see Ferguson [3] page 383 excercise 1).

Lemma 6.

If (E,P) possesses an exponentially bounded life time determined by γ and a, and ν is a real number with $1 < \nu < \gamma^{-1}$, then

$$1 \leq \sum_{n=0}^{\infty} \nu^n \mathbb{P}_{i,R}[T = n] \leq \nu^{-1} + (\nu - 1)\nu^{-1} a(i)(1 - \nu\gamma)^{-1}$$

for all $i \in E$, $R \in M$.

Proof.

$$\sum_{n=0}^{\infty} \nu^n \mathbb{P}_{i,R}[T \geq n] = \sum_{n=0}^{\infty} \nu^n \sum_{k=n}^{\infty} \mathbb{P}_{i,R}[T = k] = \sum_{k=0}^{\infty} \frac{\nu^{k+1} - 1}{\nu - 1} \mathbb{P}_{i,R}[T = k]$$

$$= \nu(\nu - 1)^{-1} \sum_{k=0}^{\infty} \nu^k \mathbb{P}_{i,R}[T = k] - (\nu - 1)^{-1} \sum_{k=0}^{\infty} \mathbb{P}_{i,R}[T = k] .$$

Hence
$$\sum_{k=0}^{\infty} \nu^k \mathbb{P}_{i,R}[T = k] = (\nu - 1)\nu^{-1} \sum_{n=0}^{\infty} \nu^n \mathbb{P}_{i,R}[T \geq n] + \nu^{-1}$$

$$\leq (\nu - 1)\nu^{-1} a(i)(1 - \nu\gamma)^{-1} + \nu^{-1} .$$ □

Theorem 2.

The life time of (E,P) is exponentially bounded iff there is on (E,P) a strongly excessive function b satisfying $0 < \delta \leq b(i)$ for all $i \in E$ and certain δ.

Proof.

We first prove the "if"-part assuming $\rho b \geq Pb$ for $P \in P$ with $0 < \rho < 1$, $b(i) \geq \delta > 0$. The assumption implies $\rho^n b \geq P_0 \ldots P_{n-1} b \geq \delta P_0 \ldots P_{n-1} e$. Hence for any $i \in E$, $R \in M$

$$\mathbb{P}_{i,R}[X_n \in E] \leq \rho^n \frac{b(i)}{\delta} .$$

So the choices $\gamma := \rho$ and $a(i) := b(i)\delta^{-1}$ prove the assertion.

We now prove the "only if"-part assuming the exponential boundedness of the life time T. We will prove that the function b defined by

$$b(i) := \sup_{R \in M} \sum_{n=0}^{\infty} \nu^n \mathbb{P}_{i,R}[T = n] , \quad i \in E$$

is strongly excessive if $1 < \nu < \gamma^{-1}$.
By lemma 6 we have $b(i) < \infty$ for $i \in E$ and therefore by lemma 5 ii) the theorem is proved. □

Corollary.

The life time of (E,P) is uniformly exponentially bounded iff (E,P) is contracting with a strongly excessive function b satisfying $0 < \delta \le b(i) \le \Delta$ for all $i \in E$ and certain δ and Δ.

Proof.

The "if"-part follows from the first part of the proof of theorem 2:

$$b(i)\delta^{-1} \le \Delta\delta^{-1}, \text{ hence } a(i) := \Delta\delta^{-1} \text{ suffices.}$$

For the "only if"-part we use the construction of the second part in the proof of theorem 2. Then we obtain b with

$$b(i) \le \nu^{-1} + (\nu - 1)\nu^{-1}a(1 - \nu\gamma)^{-1} \text{ if } a(i) = a \text{ for all } i \in E. \qquad \square$$

Remark.

In the case of Markov decision process with uniformly exponentially bounded life time, another strongly excessive function may be constructed in the following way

$$b(i) := \sup_R \sum_{n=0}^{\infty} \mathbb{P}_{i,R}[X_n \in E], \quad i \in E.$$

Namely we have $1 \le b(i) \le a(1 - \gamma)^{-1}$, $(a \ge 1)$ and

$$b(i) = \sup_R \sum_{n=0}^{\infty} \mathbb{E}_{i,R}[r(X_n)] \text{ , where } r(j) := 1 \text{ for } j \in E.$$

So by lemma 5 i) with $0 \le (1 - \rho) \le a^{-1}(1 - \gamma)$ we have that b is strongly excessive, with excessivity factor ρ.

5. Strong excessivity and the spectral radius

In this section we will present an analytical characterization of contracting Markov decision processes. If (E,P) is contracting with respect to the bounding function μ, then $\|P\|_\mu \le \rho < 1$ and consequently $\|P^n\|_\mu^{1/n} \le \rho$. So we have for a contracting Markov decision process that the spectral radii of all $P \in P$ are at most ρ:

$$\sup_{P \in P} \limsup_{n\to\infty} \|P^n\|_\mu^{1/n} \le \rho < 1 .$$

The topic of this section will be the investigation of the reverse proposition.

Definition 4.

The *spectral radius of a Markov decision process* (E,$\mathcal{P}$) *with respect to a bounding function* μ is defined as

$$\sup_{P\in\mathcal{P}} \limsup_{n\to\infty} \|P^n\|_\mu^{1/n} .$$

The main result of this section will be:

Theorem 3.

A Markov decision process (E,$\mathcal{P}$) is contracting (with respect to some bounding function) if and only if the spectral radius of (E,$\mathcal{P}$) with respect to some bounding function μ is less than one and

$$\sup_{P\in\mathcal{P}} \|P\|_\mu < \infty .$$

The "only if"-part of the theorem has been given in the introduction of this section. In the sequel of this section we will mainly be concerned with the proof of the "if"-part. In order to give the proof we will assume from now on that the spectral radius of (E,$\mathcal{P}$) with respect to the bounding function μ is equal to $\rho^* < 1$ and

$$M := \sup_{P\in\mathcal{P}} \|P\|_\mu < \infty .$$

In some steps we will prove that this Markov decision process is contracting with respect to some bounding function ν. In the first two steps we will replace the assumptions by some equivalent conditions.

Lemma 7.

There exist numbers $b_P > 0$ for all $P \in \mathcal{P}$ and $0 < \rho < 1 < \lambda$ such that

(**) $\lambda^n P^n \mu \le b_P \rho^n \mu$ for all natural n and $P \in \mathcal{P}$.

Remark.

The finiteness of M is not needed, only $\|P\|_\mu < \infty$ for all $P \in \mathcal{P}$. Furthermore, (**) implies that (E,$\mathcal{P}$) has a spectral radius smaller than one.

Proof. Choose $\lambda > 1$ with $\lambda\rho^* < 1$. Then

$$\sup_{P \in \mathcal{P}} \limsup_{n\to\infty} \| \lambda^n P^n \|_\mu^{1/n} = \lambda\rho^*$$

Choose ρ with $\lambda\rho^* < \rho < 1$. Then there is a n_P for any P with

$$\| \lambda^n P^n \|_\mu \le \rho^n \quad \text{for } n \ge n_P .$$

Hence b_P may be chosen such that

$$\| \lambda^n P^n \|_\mu \le b_P \rho^n \quad \text{for natural n.}$$ □

From now on we will use λ, ρ, b_P as the numbers satisfying (**). Actually we may suppose $\mu \equiv 1$ without loss of generality as appears from lemma 8.

Lemma 8.

Define $\mathcal{P}^*$ by $\mathcal{P}^* := \{P^* | P^*(i,j) := M^{-1}\mu^{-1}(i)P(i,j)\mu(j) \text{ , } P \in \mathcal{P}\}$. And λ^* by $\lambda^* := \lambda M$. Then

i) $\sum_j P^*(i,j) \le 1$.

ii) $\lambda^{*n} P^{*n} 1 \le b_P \rho^n$.

iii) $P^*\nu \le \lambda^{*-1}\nu$ for some bounding function ν implies $P\mu \otimes \nu \le \lambda^{-1}\mu \otimes \nu$, where $(\mu \otimes \nu)(i) := \mu(i)\nu(i)$ (and reversely).

Proof:

i) is trivial.
ii) is a consequence of the property $(P^{*n})(i,j) = M^{-n}\mu^{-1}(i)(P^n)(i,j)\mu(j)$.
iii) is proved by inspection. □

So $(E,\mathcal{P}^*)$ is a Markov decision process with spectral radius (w.r.t. $\mu \equiv 1$) smaller than one. Furthermore, lemma 8 iii) shows that for the proof of our main theorem it suffices to find a bounding function ν such that $(E,\mathcal{P}^*)$ is λ^{*-1}-contracting with respect to ν. At least, this suffices if $\lambda^* > 1$; however $\lambda^* = \lambda M > \lambda > 1$, if $M \ge 1$. For the case $M < 1$ the assertion of the theorem is trivial. So from now on, we may suppose $\mu \equiv 1$.
The proof of the theorem proceeds as follows: in lemma 9 Howard's policy iteration method is proved to converge for $(E,\mathcal{P})$ with unit rewards, discount factor λ, and finite value function. Using this method it is easy to show that $(E,\mathcal{P})$ is contracting. Then the only point which remains to be verified

is the finiteness of this value function. In fact this verification will appear to be the most tedious part of the proof.

Lemma 9.

Define v_P by $v_P := \sum_{n=0}^{\infty} \lambda^n P^n 1$ and v by $v := \sup_{P \in \mathcal{P}} v_P$. Let $v < \infty$.
Then there is a bounding function ν with $\lambda P \nu \leq \nu$ for all $P \in \mathcal{P}$ and $1 \leq \nu \leq v$.

Proof:

We start by formulating the policy iteration method for $(E,\mathcal{P})$ with reward one in each state:
Choose $P_0 \in \mathcal{P}$ and define P_{n+1} recursively for $n = 0,1,\ldots$ such that for some $\varepsilon_n(i)$ $(0 \leq \varepsilon_n(i) < 1)$

$$1 + \lambda(P_{n+1} v_{P_n})(i) \geq \sup_P\{1 + \lambda(P v_{P_n})(i)\} - \varepsilon_n(i) \geq v_{P_n}(i) .$$

The sequence $\{P_n\}$ satisfies $v_{P_n} \leq 1 + \lambda P_{n+1} v_{P_n}$, which implies

$$v_{P_n} \leq \sum_{k=0}^{N-1} \lambda^k P^k_{n+1} 1 + \lambda^N P^N_{n+1} v_{P_n} \quad \text{and hence } v_{P_n} \leq v_{P_{n+1}} .$$

Namely, $\lambda^N P^N_{n+1} v_{P_n} \leq \lambda^N P^N_{n+1} b_{P_n} (1 - \rho)^{-1} 1 \leq b_{P_{n+1}} \rho^N b_{P_n} (1 - \rho)^{-1} 1$, which tends to zero for $N \to \infty$.
Since $v_{P_n} \leq v$ and $v_{P_n} \leq v_{P_{n+1}}$ we obtain: $\nu := \lim_{n\to\infty} v_{P_n} \leq v$.
For the proof of $\nu \geq \lambda P \nu$ for all $P \in \mathcal{P}$, note that

$$v_{P_{n+1}} = 1 + \lambda P_{n+1} v_{P_{n+1}} \geq 1 + \lambda P_{n+1} v_{P_n} \geq 1 + \lambda P v_{P_n} - \varepsilon_n \text{ for all } P \in \mathcal{P}.$$

This implies $v_{P_{n+1}} \geq \lambda P v_{P_n}$, which gives $\nu \geq \lambda P \nu$. □

Remark.

It is not difficult to prove that $\nu = v$ in lemma 9, so the policy iteration method really converges. Note that we used the assumption, given in section 2, to construct the sequence $\{P_n\}$.

If we can prove that the condition $v < \infty$ is really satisfied, than the proof is completed. However in order to prove $v < \infty$ we need four more lemmas.

By P_A we will denote the sub-Markov matrix of the Markov chain with matrix P restricted to A.

Lemma 10.

Let for some $P \in \mathcal{P}$, $K \in \mathbb{R}$, $i \in E$

$$\sum_{n=0}^{\infty} \lambda^n (P^n 1)(i) \geq K \ .$$

Then there is for each $\varepsilon > 0$ a finite subset A of E with

$$\sum_{n=0}^{\infty} \lambda^n (P_A^n 1)(i) > K - \varepsilon \ .$$

(here 1 is the vector with unit components defined on A).

Proof:

(see Ornstein [9] for a similar construction).

Choose N with $\sum_{n=0}^{N} \lambda^n (P^n 1)(i) \geq v_P(i) - \frac{1}{2}\varepsilon$.

It is easy to verify that for each n there exists a finite set $A_n \subset E$ with

$$(P_{A_n}^n 1)(i) \geq (P^n 1)(i) - \frac{\varepsilon}{2N} \lambda^{-n} \ .$$

Hence $A := \bigcup_{n=0}^{N} A_n$ has the required properties. □

The foregoing lemma says that in some sense restriction to finite Markov chains is allowed for fixed P. The next lemma shows that the expected number of visits to state i (discounted with factor λ) is bounded as a function of P.

Lemma 11.

For all $P \in \mathcal{P}$ we have

$$P_i(\lambda) := \sum_{n=0}^{\infty} \lambda^n (P^n)(i,i) \leq (1 - \rho)^{-1} \qquad \text{for } i \in E \ .$$

Proof:

Fix $i \in E$ and $P \in \mathcal{P}$. Let T be the (stochastic) life time of the Markov chain with matrix P and starting state i. Let T_N be the time of the N-th recurrence to i in the same Markov chain ($T_N = \infty$ if the N-th recurrence does not occur).

For this Markov chain we have if $R = (P,P,\dots)$

$$\sum_{n=N}^{\infty} \lambda^n \mathbb{P}_{i,R}[T \geq n] \leq \sum_{n=N}^{\infty} b_P \rho^n = b_P \rho^N (1-\rho)^{-1} .$$

So $P_i(\lambda) \leq b_P(1-\rho)^{-1}$.

As usual we may express return probabilities in probabilities for the first return:

$$P_i(\lambda) = [1 - F_i(\lambda)]^{-1} = 1 + \sum_{N=1}^{\infty} F_i^N(\lambda) ,$$

with

$$F_i(\lambda) := \sum_{n=1}^{\infty} \lambda^n \mathbb{P}_{i,R}[X_n = i \ , \ X_\ell \neq i \ \text{ for } \ell = 1,\dots,n-1] = \mathbb{E}_{i,R}[\lambda^{T_1}] ,$$

where the expectation operator only indicates integration with respect to sample points with $T_1 < \infty$. With the same restriction for T_N we have as usual, using the strong Markov property:

$$\mathbb{E}_{i,R}[\lambda^{T_N}] = (\mathbb{E}_{i,R}[\lambda^{T_1}])^N = F_i^N(\lambda) .$$

Therefore

$$F_i^N(\lambda) = \mathbb{E}_{i,R}[\lambda^{T_N}] = \sum_{n=N}^{\infty} \lambda^n \mathbb{P}_{i,R}[T_N = n] \leq \sum_{n=N}^{\infty} \lambda^n \mathbb{P}_{i,R}[T \geq n] \leq \frac{b_P}{1-\rho} \rho^N .$$

This implies $F_i(\lambda) \leq \rho$ and hence $P_i(\lambda) \leq (1-\rho)^{-1}$. □

Lemma 12.

Let $B \in E$, such that for some $P \in \mathcal{P}$, and some $K \in \mathbb{R}$: $\sum_{n=0}^{\infty} \lambda^n P_B^n 1 \leq K1$.

Then we have for $B' := B \cup \{j\}$ with $j \notin B$: $\sum_{n=0}^{\infty} \lambda^n P_{B'}^n 1 \leq K'1$, with K' a constant determined by λ, ρ and K.

Proof:

We first consider starting state $i \in B$. Restrict the process to B'. Let T be the life time of the restricted process and let S be the time of the first visit to j (beofre the process leaves B', $S = \infty$ if no such visit occurs). Note that $T < \infty$ $\mathbb{P}_{i,R}$-a.s., for $R = (P,P,\dots)$.

$$\sum_{n=0}^{\infty} \lambda^n (P_B^n 1)(i) = \mathbb{E}_{i,R}[\sum_{n=0}^{\infty} \lambda^n \cdot 1_{\{T\ge n\}}]$$

$$= \mathbb{E}_{i,R}[\sum_{n=0}^{\infty} \lambda^n \cdot 1_{\{T\ge n\}} 1_{\{S=\infty\}}] + \mathbb{E}_{i,R}[\sum_{n=0}^{\infty} \lambda^n \cdot 1_{\{T\ge n\}} 1_{\{S<\infty\}}]$$

The first term is at most K by the assumption. The second will be evaluated using the strong Markov property w.r.t. S.
Note that $S \le T$ on $\{S < \infty\}$.

$$\mathbb{E}_{i,R}[\sum_{n=0}^{T} \lambda^n 1_{\{S<\infty\}}] = \mathbb{E}_{i,R}[\sum_{n=0}^{S-1} \lambda^n{}_{\{S<\infty\}} + \lambda^S 1_{\{S<\infty\}} \mathbb{E}_{j,R}[\sum_{n=0}^{T} \lambda^n]] =$$
$$\le \mathbb{E}_{i,R}[\lambda^S]\{\frac{1}{\lambda - 1} + \mathbb{E}_{j,R}[\sum_{n=0}^{T} \lambda^n]\} \le \mathbb{E}_{i,R}[\lambda^S]\{1 + \lambda \mathbb{E}_{j,R}[\lambda^T]\}(\lambda - 1)^{-1} \tag{1}$$

Note further:

$$\mathbb{E}_{i,R}[\lambda^S] = \sum_{n=0}^{\infty} \lambda^n \mathbb{P}_{i,R}[X_0 \in B,\dots,X_{n-1} \in B,\ X_n = j] \le$$
$$\le \sum_{n=1}^{\infty} \lambda^n \mathbb{P}_{i,R}[X_0 \in B,\dots,X_{n-1} \in B] \le \lambda K . \tag{2}$$

Now we consider $\mathbb{E}_{j,R}[\lambda^T]$. Let T_K be the K-th visit to j, after time zero. Again using the strong Markov property we find:

$$\mathbb{E}_{j,R}[\lambda^T] = \mathbb{E}_{j,R}[\lambda^T \cdot 1_{\{T_1=\infty\}}] + \sum_{K=1}^{\infty} \mathbb{E}_{j,R}[\lambda^T \cdot 1_{\{T_K<\infty, T_{K+1}=\infty\}}] =$$

$$= \mathbb{E}_{j,R}[\lambda^T \cdot 1_{\{T_1=\infty\}}] + \sum_{K=1}^{\infty} \mathbb{E}_{j,R}[\lambda^{T_K}] \cdot \mathbb{E}_{j,R}[\lambda^T \cdot 1_{\{T_1=\infty\}}]$$

Hence by lemma 11:

$$\mathbb{E}_{j,R}[\lambda^T] \le \mathbb{E}_{j,R}[\lambda^T \cdot 1_{\{T_1=\infty\}}]\ \{1 + \frac{1}{1-\rho}\} . \tag{3}$$

Finally we consider $\mathbb{E}_{j,R}[\lambda^{T} \cdot 1_{\{T_1=\infty\}}]$. Let $\tilde{T}$ be the life time of the process restricted to B. Then we have

$$\mathbb{E}_{j,R}[\lambda^{T} \cdot 1_{\{T_1=\infty\}}] = \lambda \sum_{i \in B} P(j,i) \mathbb{E}_{i,R}[\lambda^{\tilde{T}}] + \sum_{i \in B^c} P(j,i) \le \lambda \sup_{i \in B} \mathbb{E}_{i,R}[\lambda^{\tilde{T}}] + 1 .$$

And by the assumption of the lemma we find

$$\mathbb{E}_{j,R}[\lambda^{T} \cdot 1_{\{T_1=\infty\}}] \le (\lambda - 1)K + 2 . \tag{4}$$

Therefore we find, using (1), (2), (3) and (4):

$$\mathbb{E}_{i,R}[\sum_{n=0}^{T} \lambda^n 1_{\{S<\infty\}}] \le \lambda K[1 + \lambda(1 + \frac{1}{1-\rho})\{(\lambda - 1)K + 2\}](\lambda - 1)^{-1} .$$

For the case $i = j$ the proof follows from (3) and (4). □

Finally we prove in lemma 13, using lemmas 10 and 12, that the function v defined in lemma 9 is finite.

<u>Lemma 13.</u>

$$\sup_{i \in E} v(i) < \infty .$$

<u>Proof:</u>

Let A be a subset of E. We define on A:

$$v_A := \sup_{P \in \mathcal{P}} \sum_{n=0}^{\infty} \lambda^n P_A^n 1 .$$

Note that $v(i) = v_E(i)$, $i \in E$. We first prove that if A is finite and $\sup v_E(i) = \infty$ then $\sup_{i \in A^c} v_{A^c}(i) = \infty$. Assume the contrary: let $\sup_{i \in A^c} v_{A^c}(i) = K$. Then using lemma 12 we have for $j \in A$: $\sup_{i \in A^c \cup \{j\}} v_{A^c \cup \{j\}}(i) = K' < \infty$, and so by induction $\sup_{i \in A^c \cup A} v_{A^c \cup A}(i) < \infty$ which produces a contradiction.

Suppose $\sup_{i \in E} v(i) = \infty$. Fix a nondecreasing sequence $a_1, a_2, \ldots$ tending to infinity. Fix $\varepsilon > 0$. There must be a $i_1 \in E$ and a $P_1 \in \mathcal{P}$ such that $v_{P_1}(i_1) \ge a_1 + \varepsilon$. Hence by lemma 10 there is a finite subset A_1 such that $i_1 \in A_1$ and for $R_1 = (P_1, P_1, \ldots)$

$$\sum_{n=0}^{\infty} \lambda^n \mathbb{P}_{i_1,R_1}[X_0 \in A_1,\ldots,X_n \in A_1] \geq a_1 \ .$$

Consider the process restricted to A_1^c. We have already seen that $\sup_{i \in A_1^c} v_{A_1^c}(i) = \infty$. Hence there is a $i_2 \in A_1^c$ and a $P_2 \in P$ such that for $R_2 = (P_2,P_2,\ldots)$

$$\sum_{n=0}^{\infty} \lambda^n \mathbb{P}_{i_2,R_2}[X_0 \in A_1^c,\ldots,X_n \in A_1^c] \geq a_2 + \varepsilon$$

and again by lemma 10 there is a finite subset $A_2 \subset A_1^c$ such that $i_2 \in A_2$ and:

$$\sum_{n=0}^{\infty} \lambda^n \mathbb{P}_{i_2,R_2}[X_0 \in A_2,\ldots,X_n \in A_2] \geq a_2 \ .$$

We may apply the same argument to $(A_1 \cup A_2)^c$. So we find finite sets $A_1,A_2,\ldots$ with state $i_1 \in A_1$, $i_2 \in A_2,\ldots$ and $P_1,P_2,\ldots \in P$ such that for $R_K := (P_K,P_K,\ldots)$

$$\sum_{n=0}^{\infty} \lambda^n \mathbb{P}_{i_K,R_K}[X_0 \in A_K,\ldots,X_n \in A_K] \geq a_K \ .$$

Consider a new element $P \in P$ defined by: $P(i,j) := P_K(i,j)$ if $i \in A_K$.
It is easy to verify that for all $K = 1,2,3,\ldots$

$$v_P(i_K) \geq \sum_{n=0}^{\infty} \lambda^n \mathbb{P}_{i_K,R_K}[X_0 \in A_K,\ldots,X_n \in A_K].$$

Hence

$$\sup_K v_P(i_K) = \infty \ .$$

On the other hand we have for all $P \in P$: $v_P(i) \leq \frac{b_P}{1-\rho}$ for all $i \in E$. Therefore $\sup_i v(i) < \infty$. □

6. Some consequences and remarks.

1. In the proofs of theorem 1.2 the assumption on P has only been used for the proof of the sufficiency of both conditions for strong excessivity, not for the necessity.

2. In our definition strongly excessive functions are positive. Hinderer [5] allows the value 0 for b(i). However, the strong excessivity (even if ρ is not less than 1, as in Hinderer's case) requires for the system to remain in the set of states with b-value 0 as soon as it is entered. Hence, without restricting generality, one may assume that the state space is left when such a state is entered.

3. Combination of theorems 1 and 2 gives the following necessary and sufficient condition for exponential boundedness of the life time of (E,P): there exist a partitioning $\{E_k | k = 0,1,\ldots\}$ of E and numbers $\alpha > 1$, $\beta \geq 1$, such that for all $R \in M$

$$\sum_{n=0}^{\infty} \mathbb{P}_{i,R}[X_n \in E_k] \leq \beta \min\{1, \alpha^{\ell-k}\} \qquad \text{if } i \in E_\ell \ .$$

4. It is an immediate consequence of th. 3 that if, there are $N \in \mathbb{N}$, $M_0 \in \mathbb{R}$, $0 < \rho_0 < 1$ and a bounding function μ such that for all $P \in P$:

   i) $P\mu \leq M_0\mu$ and ii) $P^N\mu \leq \rho_0\mu$

   then the decision process is contracting.
   Hence N-stage contraction is equivalent to one-stage contraction.
   This result can also be proved straightforwardly by constructing a bounding function which makes (E,P) contracting:

$$\mu_0 := \sup_R \mathbb{E}_R \sum_{n=0}^{\infty} \alpha^{-n}\mu(X_n) \ , \quad \text{with} \quad \rho_0 < \alpha^N < 1 \ .$$

   In a decision process one considers contraction properties of a whole set P of operators simultaneously. If one only considers one (not necessarily linear) operator T in a complete matrix space X, then N-stage contraction of T implies one-stage contraction of T in X with respect to some other distance. This has been shown by Walter [12] without using the equivalent of our condition i).

5. In Veinott [11] a similarity transformation for decision processes was introduced for transient models with a finite state space. This transformation for the transition probabilities has the form

$$P^*(i,j) := \frac{P(i,j)\mu(j)}{\mu(i)} \quad P \in P, \ \mu \text{ a bounding function.}$$

   A lot of properties of the decision process are invariant under this transformation.
   Lemma 3 in Veinott's paper (due to Hoffman) is exactly the same as the statement of th. 3 for a finite state space and a finite action space. Note however that in the finite case it is obvious that the function v defined in lemma 9 is bounded. Furthermore the finiteness of E implies that $P^N\mu \leq \rho_0\mu$ for all P and some N, $\rho_0 < 1$ if the spectral radius of

(E,P) is less than one. This can easily be used to show that (E,P) is contracting (compare remark 4).

6. At first glance one may expect that

$$r := \sup_{P \in P} \sup_{i,j \in E} \limsup_{n \to \infty} \{P^{(n)}(i,j)\}^{1/n} < 1$$

is a sufficient condition for the decision process to be contracting. The quantity r may be regarded as a generalization of the concept *convergence norm* to decision processes (see Seneta [10]p. 162). However, we produce a counterexample for this statement.

<u>Counterexample.</u>

$E := \{-1,0,1,2,3,\ldots\}$, $P = \{P_n | n = 1,2,3,\ldots\}$ where for all $n = 1,2,3,\ldots$

i) $P_n(K,K-1) = 1 \qquad K \geq 1$

ii) $P_n(0,i) = 0 \quad i \neq 0$, $P_n(0,0) = \rho < 1$ and $P_n(-1,n) = 1$, $n = 1,2,3,\ldots$

It is easy to verify that if $j \neq 0$: $P_n^{(N)}(i,j) = 0$ for N sufficiently large and $\limsup_{N\to\infty} \{P_n^{(N)}(i,0)\}^{1/N} = \rho$ for all $i \in E$ and $n = 1,2,3,\ldots$. Hence $r = \rho$. However if μ is strongly excessive with excessivity factor $0 < \rho^* < 1$ then

i) $\rho^*\mu(0) \geq \rho\mu(0)$ hence $\rho^* \geq \rho$

ii) for $K = 1,2,3,\ldots$ $\rho^{*K}\mu(K) \geq \mu(0)$ hence $\mu(K) \geq \mu(0)(\rho^*)^{-K}$ and therefore since $\mu(-1) \geq \mu(K)$ $K = 0,1,2,\ldots$ we have $\mu(-1) \geq \sup_K (\rho^*)^{-K}\mu(0) = \infty$. So there does not exist a strongly excessive function here.

7. The following 6 assertions for (E,P) are equivalent:

A. (E,P) is contracting with a strongly excessive function b satisfying $0 < \delta \leq b(i) \leq \Delta$ for all $i \in E$ and certain δ and Δ.

B. (E,P) is strongly excessive with a strongly excessive function which is finitely valued.

C. For certain N and $\varepsilon > 0$

$$\mathbb{P}_{i,R}[X_N \in E] \leq 1 - \varepsilon \qquad \text{for all } i \in E,\ R \in M.$$

D. There exist a number $\varepsilon > 0$ and a finite partition $\{E_k | k = 1,\ldots,N\}$ of E, such that

$$\mathbb{P}_{i,R}[X_1 \in \bigcup_{k=\ell}^{N} E_k] \leq 1 - \varepsilon \qquad \text{for all } i \in E_\ell,\ \ell = 1,\ldots,N,\ R \in M.$$

E. The life time of (E,P) is uniformly exponentially bounded.

F. There exist a number $\rho(0 < \rho < 1)$, a function b on E with $0 < \delta < b(i) < \Delta$ for certain δ, Δ, and a natural number N, such that for all $P_1, \ldots, P_N \in P$
$P_1 \ldots P_N b \leq \rho b$.

Acknowledgement.

We are grateful to the referee for many valuable suggestions, especially for his conjecture concerning the spectral radius, which led to section 5.

References.

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

[2] E.V. Denardo, Contraction mappings in the theory underlying dynamic programming. SIAM-Review 9 (1967) 165-177.

[3] T.S. Ferguson, Mathematical Statistics, Academic Press (1967) New York and London.

[4] K.M. van Hee, Markov strategies in dynamic programming, Memorandum COSOR 75-20 (1975) submitted for publication.

[5] K. Hinderer, Bounds for stationary finite-stage dynamic programs with unbounded reward functions. Report of the "Institut für Mathematisch Stochastik der Universität Hamburg", June 1975.

[6] A. Hordijk, Dynamic programming and Markov potential theory. Amsterdam(1974)

[7] S.A. Lippman, Semi-Markov decision processes with unbounded rewards. Management Science 19 (1973) 717-731.

[8] S.A. Lippman, On dynamic programming with unbounded rewards. Management Science 21 (1975) 1225-1233.

[9] D. Ornstein, On the existence of stationary optimal strategies, Proc. Amer. Math. Soc. 20 (1969) p. 563-569.

[10] E. Seneta, Nonnegative matrices, George Allen and Unwin Ltd. (1973).

[11] A.F. Veinott, Discrete dynamic programming with sensitive discount optimality criteria, Ann. Math. Statist. 40 (1969) p. 1635-1660.

[12] W. Walter, A note on contraction, SIAM Review 18 (1976) 107-111.

[13] J. Wessels, Markov programming by successive approximations with respect to weighted supremum norms. J. Math. Anal. Appl. 58 (1977).

[14] J. Wijngaard, Stationary Markovian decision problems. Technological University Eindhoven,(1975).