Determination of the spectral radius of a Markov decision process
Zijm, W.H.M.

Published: 01/01/1979

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 author's 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

Citation for published version (APA):

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

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 14. Jan. 2019
Determinación de la raíz spectral de un proceso de decisión de Markov

by

W.H.M. Zijm

Eindhoven, Oktober 1979

The Netherlands
Determination of the spectral radius of a
Markov decision process

by
W.H.M. Zijm

0. Abstract

Consider a Markov decision process in the situation of discrete
time, finite state space and finite action space. A positive proba­bility
for fading of the system is allowed. In this case, contraction
properties of certain operators, used in Dynamic Programming, are strictly
related to the spectral radius of the process. In this paper a method
for estimating this spectral radius is proposed.

The result can be extended immediately to the case in which the
transition probability matrices are replaced by general nonnegative ma­trices.

1. Introduction

Contraction properties of certain operators in a Banach space often
play an important role in the theory of Markov decision processes. Assu­ming
a positive probability of leaving the system in N stages (uniform
in the starting state and the strategy) or, in other words, having so­
called N-step contraction, one can construct a strongly excessive func­tion, i.e. a strictly positive (column)vector $\mu$ such that for all ma­trices $P$, which are involved in the system, we have

$$(1) \quad P\mu \leq \rho \mu$$

where $\rho$, the excessivity factor, is smaller than one.

This excessivity factor can be used in determining the rate of con­vergence of a successive approximation procedure, and, closely related
to this, it can be used in estimating the value function of a Markov de­cision process (Wessels [14]). It turns out that strongly excessive
functions can be characterized in terms of the spectral radius, more
specifically: it is possible to construct strongly excessive functions
with an excessivity factor arbitrarily close to the spectral radius of
a Markov decision process (van Hee and Wessels [2], Zijm [15]), so that
an exact knowledge of this spectral radius would be very useful.

Several authors have dealt with questions concerning the spectral
radius of a Markov decision process, most of them in the irreducible case
(Isaacson and Luecke [5], Howard and Matheson [4], Mandl [7], Mandl and Seneta [8], Sladký [11], [12]). Rothblum [9] gives a Cesaro-Laurent expansion of the resolvent of a specific matrix on the spectral circle. Finally Veinott [13] presents a number of results concerning transient dynamic programming (spectral radius smaller than one).

In this paper we present a method to obtain a good approximation for the spectral radius of a Markov decision process, using the strongly excessive functions, mentioned above. In fact, we will find a sequence of strongly excessive functions \( \mu_n \) with excessivity factor \( \lambda_n \) such that the sequence \( (\lambda_n)^\infty \) is converging to the spectral radius. For that purpose we first introduce some concepts and notations and give a few preliminary results (section 2). In section 3 we develop the approximation method for the spectral radius and prove that the sequence \( (\lambda_n)^\infty \) is actually converging. In section 4 we take a closer look to the functions \( \mu_n \), and investigate their asymptotic behaviour. As a by-product of this investigation we find that all irreducible matrices with maximal spectral radius have the same, strictly positive, right eigenvector, associated with this spectral radius. Finally, in section 5, we determine upper and lower bounds for the spectral radius.

2. Preliminaries

Consider a Markov decision process with a finite state space \( S \) and a finite action space \( K \). A system is observed at discrete points of time. If at time \( t \) the state of the system is \( i \in S \) we may choose action \( k \in K \) which results in a probability \( p_{ij}^k \) of observing the system in state \( j \) at time \( t + 1 \). We suppose

\[
\sum_{j \in S} p_{ij}^k \leq 1, \text{ for all } i \in S, \ k \in K
\]

Hence a positive probability for fading of the system is allowed. A policy \( f \) is a function on \( S \) with values in \( K \). By \( F \) we denote the set of policies \( \{f | f: S \rightarrow K\} \). A strategy \( s \) is a sequence of policies:

\( s = (f_0, f_1, f_2, \ldots) \). If we use strategy \( s \), then action \( f_t(i) \) is taken, if at time \( t \) the state of the system is \( i \). A stationary strategy is a strategy \( s = (f, f, f, \ldots) \).

The spectral radius \( \rho^* \) of this Markov decision process is defined as:
(3) \[ \rho^* := \max \limsup_{n \to \infty} \| P^n(f) \|^{1/n} \]

with \( P(f) := (p_{ij})_{i,j \in S} \) and \( \| \cdot \| \) the usual sup-norm.

Throughout this paper we assume \( \rho^* < 1 \).

We say that a vector \( v \) is nonnegative (positive) - written \( v \geq 0 \) (\( v \gg 0 \)) - if all its components are nonnegative (positive). We say that \( v \) is semipositive - written \( v > 0 \) - if \( v \geq 0 \) and \( v \neq 0 \). We write \( v \geq w \) (\( v \gg w, v > w \)), if \( v - w \geq 0 \) (\( v - w \gg 0, v - w > 0 \)). By \( e \) we denote the vector with all components equal to one, by I the identity matrix.

For \( \lambda > \rho^* \) we define:

(4) \[ z(\lambda) := \max_{f} (I - P(f))^{-1} e = \max_{f} \sum_{n=0}^{\infty} \lambda^{-(n+1)} P^n(f)e \]

(max taken component - wise)

This concept will play a central role in the approximation method. The following properties can be proved:

Lemma 1: For \( \lambda > \rho^* \) it holds:

a) \( \| z(\lambda) \| < \infty \)

b) \( z(\lambda) = \sup_{s} \sum_{n=0}^{\infty} \lambda^{-(n+1)} \sup_{t=0} P(f^s_t)e \), where \( s = (f_0, f_1, f_2, \ldots) \)

c) \( \lambda z(\lambda) = \max_{f} \{ e + P(f)z(\lambda) \} \)

Proof: a) follows from a more general result in van Hee and Wessels [2].

In this case it is immediate from

\[ \max \limsup_{n \to \infty} \| \lambda^{-(n+1)} P^n(f) \|^{1/n} = \lambda^{-(1-\rho^*)} < 1. \]

The proof of b) and c) can be found in Zijm [15].

From lemma 1c, it follows that, for \( \lambda > \rho^* \), we have

(5) \[ P(f)z(\lambda) \leq \lambda z(\lambda), \quad \text{for all } f \in F \]

For \( \lambda < 1 \) \( z(\lambda) \) is an example of a strongly excessive function.

Definition: A strongly excessive function of a Markov decision process is a (column)vector \( \mu \gg 0 \), such that for some \( \rho < 1 \) (the excessivity factor) we have

(6) \[ P(f)\mu \leq \rho \mu , \quad \text{for all } f \in F \]
**Lemma 2:** For a Markov decision process with spectral radius $\rho^* < 1$ and strongly excessive function $\mu$ with excessivity factor $\rho$ we have $\rho \geq \rho^*$.

**Proof:** Choose constants $c_1, c_2$ such that $c_1 e \leq \mu \leq c_2 e$. Then

$$\rho^* = \max_{f} \limsup_n \| c_1 c_2^{-1} p^n(f) \| n \leq \rho.$$  

**Remark:** Strongly excessive functions in the case of a countable state space and general action space are treated extensively in van Hee and Wessels [2].

We see that $\rho^*$ is a lower bound for the excessivity factor of any strongly excessive function. Moreover, (5) shows that it is possible, e.g. by a policy iteration procedure ([3]), to construct strongly excessive functions with an excessivity factor arbitrarily close to the spectral radius $\rho^*$, if the last one is exactly known (Choose $\lambda$ such that $\rho^* < \lambda \leq \rho^* + \varepsilon$, $\varepsilon$ small).

In the following, however, we suppose $\rho^* < 1$, but otherwise unknown, and we use the concept of strongly excessive functions to find a good estimate for it.

### 3. Approximation of the spectral radius

In this section we will construct a sequence of numbers $(\lambda_n)_{n=0}^{\infty}$ and a sequence of vectors $(z(\lambda_n))_{n=0}^{\infty}$ (defined by (4)) such that

$$\lim_{n \to \infty} \lambda_n = \rho^*,$$

where $z(\lambda_n)$ is needed for the calculation of $\lambda_{n+1}$. The basic idea of the method is as follows:

Consider, for $\lambda \geq \rho^*$ the functional equation

$$\lambda z = \max_{f} \{ e + P(f)z \}.$$
From (4) and lemma 1 it follows that the unique, strictly positive finite solution \( z = z(\lambda) \) exists if and only if \( \lambda > \rho^* \). Hence the existence of such a solution is a method to determine whether \( \lambda > \rho^* \) or \( \lambda = \rho^* \). Moreover, if \( \lambda > \rho^* \), then

\[
\max_f P(f)z(\lambda) \leq \lambda z(\lambda) - \epsilon \leq \lambda(1 - \frac{1}{\lambda \|z(\lambda)\|})z(\lambda)
\]

Hence, by lemma 2,

\[
\rho^* \leq \lambda(1 - \frac{1}{\lambda \|z(\lambda)\|}) < \lambda
\]

and we may investigate (8) again, now for \( \lambda' := \lambda(1 - \frac{1}{\lambda \|z(\lambda)\|}) \).

In general, starting with some \( \lambda_0 > \rho^* \), e.g. \( \lambda_0 = 1 \), we define for \( n \geq 0 \):

1) \( \lambda_{n+1} = \lambda_n(1 - \frac{1}{\lambda_n \|z(\lambda_n)\|}) \), if (8) has a strictly positive, finite solution \( z = z(\lambda_n) \), for \( \lambda = \lambda_n \).

2) \( \lambda_{n+1} = \lambda_n' \), otherwise

The convergence of \( (\lambda_n)_{n=0}^\infty \) is established in the following theorem.

**Theorem 1:** \( \lim_{n \to \infty} \lambda_n = \rho^* \)

**Proof:** It is immediately clear that \( \lambda_n \geq \rho^* \), \( n = 0,1,2,... \), which implies (cf. lemma 1) that we define \( \lambda_{n+1} = \lambda_n \) if and only if \( \lambda_n = \rho^* \). Hence we may suppose that \( (\lambda_n)_{n=0}^\infty \) is strictly decreasing. Suppose

\( \gamma := \lim_{n \to \infty} \lambda_n > \rho^* \) (hence \( \gamma > 0 \))

Then \( \|z(\gamma)\| < \infty \), and, since \( \lambda z(\lambda) \) is decreasing in \( \lambda \),

\[
(1 - \frac{1}{\lambda_n \|z(\lambda_n)\|}) \leq (1 - \frac{1}{\gamma \|z(\gamma)\|}) < 1 \quad \text{for } n = 0,1,2,...
\]
which implies:
\[ \lim_{n \to \infty} \lambda_n \leq \lim_{n \to \infty} \left(1 - \frac{1}{\gamma\|z(\gamma)\|}\right)^n \lambda_0 = 0. \]

Because of the contradiction, we must conclude
\[ \lim_{n \to \infty} \lambda_n = \rho^* \]

In order to avoid trivialities we suppose in the rest of this paper that always the first possibility occurs, i.e. that we will find a monotone decreasing sequence \((\lambda_n)_{n=0}^{\infty}\) such that \(\lambda_n > \rho^*, \) for all \(n, \) and
\[ \lim_{n \to \infty} \lambda_n = \rho^* \]

In the following section we will take a closer look to the asymptotic behaviour of the sequence \(\{z(\lambda_n)\}_{n=0}^{\infty}\), for two reasons. First of all, the study of this asymptotic behaviour leads to a few nice properties of the Markov decision process. Secondly, we will need the results, in order to derive lower bounds for the spectral radius.

4. Asymptotic behaviour of \(\{z(\lambda_n)\}_{n=0}^{\infty}\)

In this section we study the asymptotic behaviour of \(\{z(\lambda_n)\}_{n=0}^{\infty}\).

More specifically, defining
\[ \mu_n = \frac{z(\lambda_n)}{\|z(\lambda_n)\|} \] (11)
we will show the existence of a vector \(\mu > 0\), such that
\[ \lim_{n \to \infty} \mu_n = \mu \] (12)

We will need some statements from the theory of linear operators on a finite dimensional vector space (Dunford and Schwartz [1], Kato [6]).

The resolvent \(R(\lambda, P)\) of an \(S \times S\) -matrix \(P\) is defined as
\[ R(\lambda, P) = (\lambda I - P)^{-1}, \quad \text{for} \ \lambda \in \mathbb{C} \setminus \sigma(P) \] (13)
Here \( \sigma(P) \) denotes the spectrum of \( P \). The spectral radius of \( P \) is denoted by \( \rho(P) \); recall that for nonnegative matrices \( \rho(P) \) is real and nonnegative, that \( \rho(P) \) equals the largest eigenvalue of \( P \), and that we may choose the corresponding left- and right-eigenvector nonnegative (Perron-Frobenius theorem, see e.g. Seneta [10]).

From Kato [6] it is seen that we may give a Laurent expansion of \( R(\lambda, P) \) at \( \lambda = \rho(P) \), for \( \lambda \) sufficient close to \( \rho(P) \), which takes the form:

\[
R(\lambda, P) = \sum_{n=-k(P)}^{\infty} (\lambda - \rho(P))^n A_n(P)
\]

We will not consider in detail the meaning of the matrices \( A_n(P) \) and of \( k(P) \) (see for example Rothblum [9]), we only note that \( k(P) \leq S \). It follows that \( \rho(P) \) is a pole of \( R(\lambda, P) \) of order \( k(P) \).

In the same way we may write:

\[
R(\lambda, P)e = \sum_{n=-k(P)}^{\infty} (\lambda - \rho(P))^n e_n(P)
\]

where \( e_n(P) := A_n(P)e \), for \( n = -k(P), -k(P) + 1, \ldots \)

Furthermore,

\[
e^{-k(P)}(P) = \|e^{-k(P)}(P)\| \cdot \lim_{\lambda \uparrow \rho(P)} \frac{R(\lambda, P)e}{\|R(\lambda, P)e\|} > 0
\]

Returning to our notations of the preceding sections, notice that for \( \lambda > \rho^* \):

\[
z(\lambda) = \max(\lambda I - P(f))^{-1}e = \max_{f} R(\lambda, P(f))e
\]

Let

\[
F_0 := \{ f : S \to K \mid \rho(P(f)) < \rho^* \}
\]
and

\[ \bar{\rho} = \max_{f \in F_0} \rho(P(f)) \]

Since \( F_0 \) is finite, there exists a constant \( C \), such that for \( \lambda \geq \rho^* > \bar{\rho} \)

\[
\max_{f \in F_0} \sum_{n=0}^{\infty} \lambda^{-(n+1)} P^n(f) e \leq C \cdot e
\]

However:

\[
\lim_{\lambda \to \rho^*} \| z(\lambda) \| = \infty.
\]

Returning to the sequence \( \{\lambda_n\}_{n=0}^{\infty} \) of section 3 this means that for \( n \) large enough, we are only dealing with policies \( f \) such that \( \rho(P(f)) = \rho^* \). Combining (15) and (17), we may write for \( \lambda_n \) close enough to \( \rho^* \) (hence for \( n \) large enough):

\[
z(\lambda_n) = \max_{f \in F_1} \sum_{l=-k(P(f))}^{\infty} (\lambda_n - \rho^*)^l e_{l}(P(f))
\]

with

\[ F_1 := \{ f \mid \rho(P(f)) = \rho^* \}, \text{ hence } F_1 = F \setminus F_0 \]

Defining:

\[ k(1) = \max_{f \in F_1} k(P(f)) \]

and

\[ e_{-k(1)} := \max_{f \in F} e_{-k(1)}(P(f)) \quad (e_{-k(1)}(f) = 0 \text{ if } k(P(f)) < k(1)) \]

we conclude from (20) and (16)

\[
\lim_{n \to \infty} \mu_n = \lim_{n \to \infty} \frac{z(\lambda_n)}{\| z(\lambda_n) \|} = \frac{e_{-k(1)}}{\| e_{-k(1)} \|} > 0.
\]

We establish to immediate corollaries of the existence of \( \lim_{n \to \infty} \mu_n \).
Theorem 2: There exists a vector \( \mu > 0 \) such that
\[
\max_{f} P(f)\mu = \rho^* \mu
\]

Proof: From lemma 1c and (11) we have
\[
\lambda_n \mu_n = \max_{f} \left\{ \frac{e}{\|z(\lambda_n)\|} + P(f)\mu_n \right\}
\]
Define \( \mu := \frac{e^{-k(1)}}{\|e^{-k(1)}\|} \), then, by (21) and theorem 1, the result follows.

For the next result we need the concept of an incidence matrix.

Definition: The incidence matrix of a Markov decision process is a matrix \( P \), defined by
\[
P = \{ p_{ij} \mid i,j \in S; p_{ij} = 1 \text{ if } p^{f(i)} > 0 \text{ for some } f, \]
\[
p_{ij} = 0 \text{ otherwise} \}
\]

Theorem 3: If the incidence matrix of a Markov decision process is irreducible (we say that the system is communicating) then the vector \( \mu \), defined in theorem 2, is strictly positive.

Proof: Suppose \( \mu(i) = 0 \), for \( i \in D \subseteq S \). Then, since \( P(f)\mu \leq \rho^* \mu \), for all \( f \), we have
\[
p^{f(i)} = 0, \text{ for } i \notin D, j \in D; \text{ for all } f \in F.
\]
Hence, for the incidence matrix \( P \):
\[
p_{ij} = 0, \text{ for } i \notin D, j \in D
\]
which is impossible, since \( P \) is irreducible. Hence \( D = \emptyset \).
Corollary: In a communicating system (irreducible incidence matrix) each irreducible transition matrix with maximal spectral radius \( \rho^* \), possesses exactly the same, strictly positive, right eigenvector \( \mu \), associated with \( \rho^* \).

Proof: Let \( P(f) \) be irreducible, and \( \rho(P(f)) = \rho^* \). According to the well-known Perron-Frobenius theorem, \( P(f) \) possesses strictly positive left- and right eigenvectors, associated with \( \rho^* \). From \( P(f)\mu \leq \rho^* \mu \) it follows, by multiplying with the left eigenvector, that \( P(f)\mu = \rho^* \mu \), where \( \mu > 0 \) (theorem 3).

In the final section we will use the fact that \( \lim_{n \to \infty} \mu_n \) exists to determine upper and lower bounds for the spectral radius at each step of the approximation procedure, and we prove the convergence of these bounds to \( \rho^* \). At the end we repeat the whole procedure.

5. Upper and lower bounds for the spectral radius

In this section we will determine upper and lower bounds for the spectral radius at each step of the approximation procedure. Obviously \( \lambda_n, n = 0, 1, 2, \ldots \) may serve as an upper bound for \( \rho^* \). In order to determine a lower bound we define

\[
\alpha_n = \min_{f, i} \frac{\max P(f)z(\lambda_n)(i)}{(z(\lambda_n))(i)}
\]

Then the following result holds

Lemma 3: \( \alpha_n \leq \rho^* \)

Proof: From (11) and (22) we conclude that there exists a policy \( \bar{f} \), such that

\[
P(\bar{f})\mu_n = \max_f P(f)\mu_n \geq \alpha_n \mu_n \gg 0
\]
Completely analogous to the proof of lemma 2 we find that
\[ \limsup_{k \to \infty} \| P^k(x) \| k \geq a_n \]
hence
\[ \rho^* \geq a_n \]

**Theorem 4:** If the system is communicating, then
\[ \lim_{n \to \infty} a_n = \rho^* \]

**Proof:**
\[ \lim_{n \to \infty} a_n = \lim_{n \to \infty} \min_i \frac{f}{\mu_n(i)} = \min_i \frac{f}{\mu(i)} = \rho^* \]

since \( \mu \) exists (theorem 2 and (21)) and \( \mu > 0 \). (theorem 3)

Hence in the communicating case we have converging upper and lower bounds (theorem 1 and theorem 4). The following simple example shows that difficulties may arise in the general case.

**Example:** Suppose we have a system with two states, 1 and 2, and only one policy \( f \). Let
\[ P(f) = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} \]

Hence \( \rho^* = \frac{3}{4} \), \( \mu = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \), while for \( n \) large enough \( \mu_n \) will take the form:
\[ \mu_n = \begin{pmatrix} \epsilon_n \\ 1 \end{pmatrix}, \text{ with } \lim_{n \to \infty} \epsilon_n = 0 \]

By (22) \[ \lim_{n \to \infty} a_n = \frac{1}{2} < \rho^*. \]
To avoid the difficulties, illustrated above, we proceed as follows.
In the case that the incidence matrix $P$ is reducible we may, eventually after permuting rows and corresponding columns, write

$$
P = \begin{pmatrix}
  P_{11} & P_{12} \\
  P_{21} & P_{22} \\
  \cdot & \cdot \\
  P_{r1} & \cdots & P_{rr}
\end{pmatrix}
$$

where $P_{11}, 1 = 1, \ldots, r$ is irreducible and $P_{1k} = 0$ for $k > 1$. Denote by $C_1$ the subset of states, corresponding to $P_{11}$, for $l = 1, \ldots, r$. Now, for all $f \in F$, define the matrix $Q(f) := (q_{ij}^{f(i)})_{i,j \in S}$ by

$$
q_{ij}^{f(i)} = \begin{cases}
  f(i)P_{ij}, & \text{if } i,j \in C_1, \text{ for some } l \in \{1, \ldots, r\} \\
  0, & \text{elsewhere}
\end{cases}
$$

It is well-known that $Q(f)$ possesses the same eigenvalues as $P(f)$, for $f \in F$, in particular $\rho^* = \max_f \rho(Q(f))$. Let $Q_{11}(f) := P_{11}(f)$ and define, for $l = 1, \ldots, r$.

$$
\rho^*_l := \max_f \rho(Q_{11}(f))
$$

then it is easy to see that we can estimate $\rho^*_l, 1 = 1, \ldots, r$, simply by applying the procedure, described in section 3, to $r$ different problems now. If we restrict ourselves to the Markov decision process with transition matrices $Q_{11}(f), f \in F$, then we have a communicating system, and hence, by theorem 1 and 4, we can determine upper and lower bounds, which are converging to $\rho^*_l$. If we are only interested in $\rho^*$ then the procedure may be the following:
Step 0: Calculate the incidence matrix $P$ of the Markov decision process. Determine its irreducible classes, $C_1,\ldots,C_x$ say, and cancel, for all matrices $P(f)$, the transition between $C_k$ and $C_l$, $k,l = 1,\ldots,x$; $k \neq l$.

Step 1: Choose $\delta > 0$, define $\lambda_0 = 1$.

For $n = 0,1,2,\ldots$

Step 2: Investigate the functional equation

\[ \lambda_n z = \max_f \{ e + P(f)z \} \]

If (26) has no strictly positive finite solution, then go to A.

Otherwise, define

\[ \lambda_{n+1} = \lambda_n \left( 1 - \frac{1}{\lambda_n \|z(\lambda_n)\|} \right) \]

Step 3: Calculate

\[ \alpha_n = \max_{l=1,\ldots,x} \min_{i \in C_l} \frac{\max_f P(f)z(\lambda_n)(i)}{z(\lambda_n)(i)} \]

If $\lambda_{n+1} - \alpha_n \leq \delta$ then go to B

Otherwise, increase $n$ with 1 and return to Step 2

A: $\lambda_n = \rho^*$; go to Stop

B: $\alpha_n \leq \rho^* \leq \lambda_{n+1}$; go to Stop

Stop.

We will end this paper with some remarks. First of all, the proofs in this paper remain unchanged if we replace the substochastic matrices by general nonnegative matrices. The only difference is that we must choose $\lambda_0$ such that $\lambda_0 > \rho^*$ (for instance, let $\lambda_0$ be the maximal row sum, where the maximum is taken over all rows and all matrices).

In every step of the approximation procedure we have to solve a set of functional equations. This can be done e.g. by using Howard's policy iteration procedure (see Howard [3]). However, for a large state space
the amount of work will grow rapidly. For that reason we may use approximations $z'(\lambda_n)$, in stead of $z(\lambda_n)$, in order to improve $\lambda_n$. Using a specific type of approximations, again convergence of the sequence $(\lambda_n)_{n=0}^{\infty}$ to $\rho^*$ can be proved (see Zijm [15]).

Finally, notice that the idea of the investigation of the functional equation (8) has also been used in Mandl [7], although furthermore the approximation technique is quite different. Moreover, in [7] only Markov decision processes with strictly positive transition matrices are studied.

Acknowledgement: I am grateful to Jan van der Wal and to Professor Uriel G. Rothblum for several valuable suggestions, especially with respect to section 4.

References


