Bounding functions for Markov decision processes in relation to the spectral radius
Zijm, W.H.M.

Published: 01/01/1978

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. Dec. 2018
EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics

PROBABILITY THEORY, STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 78-21

Bounding functions for Markov decision processes in relation to the spectral radius

by

W.H.M. Zijm

Eindhoven, October 1978

The Netherlands
Bounding functions for Markov decision processes
in Relation to the spectral radius
by
W.H.M. Zijm

Abstract

We consider Markov decision processes in the situation of discrete
time, countable state space and general decision space. By introducing
so-called "weighted supremum norms" or "bounding functions", conver­
gence of successive approximations to the value function can be proved
under certain conditions. These bounding functions may also be
applied to reduce the norms of the transition probability matrices
and hence (mostly) to improve upper and lower bounds of the approxima­
tion procedure.
In this paper we will show, that it is possible to construct bounding
functions which are strongly excessive with an excessivity factor
arbitrarily close to the spectral radius of the Markov decision
process, where this spectral radius is assumed to be smaller than one.

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 systems in N stages
(uniform in the starting state and the strategy), or, in other words,
having so-called N-step contraction, we can construct a strongly
excessive function, i.e. an excessive function with an excessivity
factor smaller than one. With the help of this strongly excessive
function we can construct a norm, the weighted supremum norm (see
Wessels [7]), so that we have 1-step contraction in this norm with a
contraction factor equal to the excessivity factor. This contraction
factor plays an important role in the estimates for the value function
of a Markov decision process and one may ask how good, that means, how small we can get it. In other words, can we construct strongly excessive functions with an excessivity factor as small as possible in order to get a good estimation for the value function, and under which conditions?

An important result has been stated in a paper by van Hee and Wessels [2], where a strongly excessive function was proved to exist with an excessivity factor $\rho^* + \varepsilon$, where $\rho^*$ is defined as the spectral radius of a Markov decision process and supposed to be smaller than one, and $\varepsilon > 0$, arbitrarily small. In this paper the same result will be proved, but the method is essentially simpler and more constructive. Where [2] uses a policy iteration method and takes the limit as a strongly excessive function, we work with a successive approximation procedure and use the result, provided by the n-th step, i.e. we may stop after a finite number of iterations. Before doing so, we will first state some preliminary results of which especially lemma 1 and lemma 3 are worth mentioning. Lemma 1, due to van Hee and Wessels [2], states that, having a Markov decision process with a spectral radius smaller than one, the supremum, taken over all stationary Markov strategies, of the lifetime of the process will be finite. Lemma 3 gives the natural extension to all (non-randomized) Markov strategies.

In this paper, the definitions and notations of Wessels [7] will be used. We consider a Markov decision process with a countably infinite state space $S$ and general decision space $K$. A system is observed at discrete points of time $(t = 0, 1, 2, \ldots)$. If at time $t$ the state of the system is $i \in S$, we may choose decision $k \in K$, which results in a reward $r_{ik}^k$ and a probability $P_{ij}^k$ of observing the system in state $j$ at time $t + 1$. We suppose

$$\sum_{j \in S} P_{ij}^k \leq 1 \text{ for all } i \in S, k \in K$$

A policy $f$ is a function on $S$ with values in $K$. A strategy $s$ is a sequence of policies: $s = (f_0, f_1, f_2, \ldots)$. If strategy $s$ is used we take decision $f_t(i)$, if a time $t$ the state of the system is $i$. A stationary strategy is a strategy $s = (f, f, f, \ldots)$. 
As optimality criterion we choose the total expected reward, which is defined for a strategy \( s = (f_0, f_1, f_2, \ldots) \) as a vector \( V(s) \) in the following way:

\[
(1) \quad V(s) = \sum_{t=0}^{\omega} \left\{ \prod_{n=0}^{t-1} P(f_n) \right\} r(f_t)
\]

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 \( r_1^f(i) \) (for \( i \in S \)) for any policy \( f \) and \( P(f) \) is interpreted as a matrix with \((i,j)\)-component \( p_{ij}^f \) (for \( i,j \in S \)) for any policy \( f \).

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

\[
(2) \quad \sum_{j \in S} p_{ij}^f \leq \rho < 1, \quad \left| r_1^f(i) \right| \leq M_0 \quad (\text{for all } i \in S, k \in K)
\]

Under these conditions the total expected reward \( V(s) \) is at most \( M_0 / (1 - \rho) \). 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 \( s = (f, f, \ldots) \) with \( V(s) \) lying between the same bounds. ([3], [4], [5], [1]).

Wessels [7] treated a more general situation. Instead of (2) he only assumes the existence of a positive function \( \mu \), defined on \( S \), such that

\[
(3) \quad \sup_i \mu^{-1}(i) \left| r_1^f(i) \right| \leq M_0 < \infty, \text{ and}
\]

\[
(4) \quad \sup_k \sup_i \mu^{-1}(i) \sum_{j \in S} p_{ij}^f \mu(j) =: \rho_H < 1
\]

Remark: \( \mu \) is a strongly excessive function, with excessivity factor \( \rho_H \), i.e. \( P(f) \mu \leq \rho_H \mu \), \( \forall f \).

Formules (3) and (4) imply a norm, the weighted supremum norm,
namely for vectors \(v\):

\[
\|v\| := \sup_{i} \mu^{-1}(i) |v(i)|, \text{ for all } v, \text{ for which } \sup_{i} \mu^{-1}(i) |v(i)| < \infty.
\]

and for matrices \(A\).

\[
\|A\| := \sup_{i} \mu^{-1}(i) \sum_{j \in S} |a_{ij}| \mu(j)
\]

Introduce (5) \(L_f v := x(f) + P(f)v\)

(6) \(V_f := \sup_{f} L_f v\)

Wessels [7] then proves, that the following successive approximation procedure ends after a finite number of steps:

**Start**

Choose \(\alpha > 0, \delta > 0, v_0 \) with \(\|v_0\| \mu < \infty\) and \(v_0 < Uv_0\) (for all components), and \(\delta (1 - \rho_H)^{-1} < \alpha\)

**Iteration part**

Find for \(n = 1, 2, \ldots\) a policy \(f_n\) such that

\[ v_n := L_{f_n} v_{n-1} \geq \max(v_{n-1}, Uv_{n-1} - \delta \mu) \]

until

\[
\frac{\delta + \rho_H \sup_{i} \mu^{-1}(i)(v_n(i) - v_{n-1}(i))}{1 - \rho_H} - \frac{\rho_L \inf_{i} \mu^{-1}(i)(v_n(i) - v_{n-1}(i))}{1 - \rho_L} < \alpha
\]

where \(\rho_L := \inf_{i} \mu^{-1}(i) \sum_{j} P_{ij} \mu(j)\)

**Stop**

Then:

\[
V_n + \frac{\rho_L}{1 - \rho_L} \inf_{i} \mu^{-1}(i)(v_n(i) - v_{n-1}(i)) \leq V(f_n) \leq V \leq V_n + \frac{\rho_H}{1 - \rho_H} [\delta + \sup_{i} \mu^{-1}(i)(v_n(i) - v_{n-1}(i))]\]

The important question now is: can we choose an appropriate strongly excessive function \(\mu\), or, once having a \(\mu\), can we improve it in order to obtain better bounds for the value function \(V\)? Therefore we will
5

give a characterization of strongly excessive functions in terms of the spectral radius of a Markov decision process.

2. Preliminaries

Suppose that we have a Markov decision process and a positive function $\mu$ such that for all policies $f$ the following holds:

\begin{equation}
\tag{9} P(f)\mu \leq M\mu
\end{equation}

for some positive constant $M$.

(9) implies

\[ P^n(f)\mu \leq M^n\mu. \]

Hence

\[ \|P^n(f)\|_\mu \leq M \]

and therefore:

\begin{equation}
\tag{10} \rho^* := \sup_{f} \limsup_{n \to \infty} \|P^n(f)\|_\mu \leq M
\end{equation}

So $\rho^*$ (which is called the spectral radius of the Markov decision process) appears to be a lower bound for $M$.

Notice that for a single finite matrix its spectral radius is equal to the absolute value of the largest eigenvalue. This explains why $\rho^*$ does not depend on $\mu$ in that case. For a single infinite matrix, i.e. $S$ countably infinite, it is easy to see that for two positive functions $\mu$ and $\tilde{\mu}$ with $\alpha\mu \leq \tilde{\mu} \leq \beta\mu$ ($\alpha, \beta$ positive constants)

\[ \limsup_{n \to \infty} \|P^n(f)\|_\mu^{1/n} = \limsup_{n \to \infty} \|P^n(f)\|_{\tilde{\mu}}^{1/n}. \]

Hence also in a Markov decision process with countably infinite state space the spectral radius is the same for $\mu$ and $\tilde{\mu}$.

Now, the main topic of this paper will be to prove, that, having $\rho^* < 1$, it is possible to construct a strongly excessive function with an excessivity factor almost equal to $\rho^*$. 
We first formulate our assumptions. Suppose we have a bounding function \( \mu \) such that:

\[
\begin{align*}
(11) \quad \rho^* &= \sup_{f} \limsup_{n \to \infty} \| P^n(f) \|^n \mu < 1, \text{ and} \\
(12) \quad M &= \max \{1, \sup_{f} \| P(f) \| \mu \} < \infty
\end{align*}
\]

Furthermore, choose \( \lambda \) with \( 1 \leq \lambda < \rho^{*-1} \).

In order to simplify the proofs and notations of the following lemmas we next give a transformation of the Markov decision process (see van Hee and Wessels [2], lemma 6, or Veinott [6], lemma 3).

Define \( P^*(f) \) as the matrix with elements \( P_{ij}^*(f) \), \( i, j \in S \), where

\[
\begin{align*}
(13) \quad P_{ij}^*(f) &= M^{-1} \mu^{-1}(i) P_{ij} f(i) \mu(j) \\
\end{align*}
\]

and \( \lambda^* \) by:

\[
(14) \quad \lambda^* = \lambda M
\]

Then, assumptions (11) and (12) become:

\[
\begin{align*}
(15) \quad \sup_{f} \limsup_{n \to \infty} \| P^n(f) \|^n = M^{-1} \rho^* \quad (\text{where } \| \cdot \| \text{ denotes the usual sup-norm}) \quad \text{and hence} \\
\sup_{f} \limsup_{n \to \infty} \| \lambda^* P^n(f) \|^n = \lambda^* M^{-1} \rho^* = \lambda^* \rho^* < 1
\end{align*}
\]

\[
(16) \quad P^*(f) \in \mathbb{E} (\text{where } \mathbb{E} \text{ denotes the unit vector})
\]

This transformation enables us to work with the usual sup-norm, instead of the weighted supremum norm. For the rest of this section we will write \( \lambda \) and \( P(f) \) again, instead of \( \lambda^* \) and \( P^*(f) \).

Hence we may write:

\[
\begin{align*}
(17) \quad \sup_{f} \limsup_{n \to \infty} \| \lambda^n P^n(f) \|^n &= 1
\end{align*}
\]
Lemma 1: Define \( z_f := \sum_{n=0}^{\infty} \lambda^n P^n(f) e \) and \( z := \sup_f z_f \) (where the sup is taken component wise). Let furthermore (17) and (18) hold. Then \( \|z\| < \infty \).

For a proof, see van Hee and Wessels [2], lemma 11.

Lemma 2: For \( z \) defined in lemma 1, and under conditions (17), (18) we have: \( z \geq e + \lambda P(f) z, \forall f \)

Comment: This result is not trivial. If \( z \) was defined as the supremum, taken over all (non-randomized) Markov strategies, then \( z \) would satisfy Bellman's optimality principle and the above inequality would be obvious. However \( z \) is only defined as the supremum, taken over all stationary Markov strategies.

Proof: Suppose that for \( i \in S, k \in K \) we have

\[
z(i) < 1 + \lambda \sum_{j \in S} p_{ij}^k z(j) \text{ or } z(i) + \varepsilon = 1 + \lambda \sum_{j \in S} p_{ij}^k z(j), \quad \varepsilon > 0
\]

Choose \( \delta = \frac{\varepsilon}{2 \|z\|} > 0 \) (see lemma 1)

Because of the definition of \( z \) we have, for all \( a \in S \):

\[
\exists f_a : S \to K \text{ such that } z(a) \geq \sum_{n=0}^{\infty} \lambda^n \sum_{b \in S} (P^n(f_a)_{a,b} \geq z(a) - \delta. \text{ Hence } \]

\[
1 + \lambda \sum_{j \in S} p_{aj}^f(a) \{ \sum_{n=0}^{\infty} \lambda^n \sum_{b \in S} (P^n(f)_{j,b} \geq z(a) - \delta
\]

and so:

\[
1 + \lambda \sum_{j \in S} p_{aj}^f(a) z(j) \geq z(a) - \delta
\]
Define $f$ as follows:

$$f(i) := k$$

$$f(a) := f(a) \text{ for } a \neq i$$

Then we have

$$z(i) - \delta + \varepsilon < z(i) + \varepsilon = 1 + \lambda \sum_{j \in S} f(i) z(j)$$

$$z(a) - \delta \leq 1 + \lambda \sum_{j \in S} f(a) z(j)$$

Define $\varepsilon_i$ as the (column) vector with $i$-component equal to $\varepsilon$, and all other components equal to zero.

Then: $z - \delta e + \varepsilon_i \leq e + \lambda P(f)z$

or $e \geq z - \lambda P(f)z + \varepsilon_i - \delta e$

and so:

$$\sum_{n=0}^{N} \lambda^n P^n(f) e \geq z - \lambda^{N+1} P^{N+1}(f) z + \sum_{n=0}^{N} \lambda^n P^n(f) \varepsilon_i - \delta z \geq z - \lambda^{N+1} P^{N+1}(f) \| z \| e + \varepsilon_i - \delta \| z \| e \quad \forall n \in \mathbb{N}.$$ 

For $N \rightarrow \infty$ we have

$$e \geq \lim_{N \rightarrow \infty} \sum_{n=0}^{N} \lambda^n P^n(f) e \geq z + \varepsilon_i - \delta \| z \| e = z + \varepsilon_i - \frac{1}{2} \varepsilon_i e$$

(Remark: For $N$ sufficiently large we have $\| \lambda^{N+1} P^{N+1}(f) \| < \rho^{N+1}$, where $\rho$ is taken in such a way that $\sup_{f} \limsup_{n \rightarrow \infty} \| \lambda^n P^n(f) \|^{1/n} < \rho < 1$

Hence: $z(i) \geq z(i) + \frac{1}{2} \varepsilon_i$ which proves the contradiction. q.e.d.

Lemma 3. Let $z$ be defined as in lemma 1 and let (17) and (18) hold.

Then (19) $z = \sup_{s \in \mathbb{N}} \sum_{n=0}^{\infty} \lambda^n P^s e$ where $s = (f_0, f_1, f_2, \ldots)$

Proof: follows immediately from lemma 2.

Once having lemma 3, we know that the following successive iteration procedure is converging:
\[ z_0 \equiv 0 \]
\[ z_n = \sup_{f} \{ e + \lambda P(f) z_{n-1} \}; \quad n = 1, 2, \ldots \]

(Note that \( z_n \leq z, \forall n \), and that \( (z_n)_{n=1}^{\infty} \) is an increasing sequence).

The next lemma states that \( z_n \) converges to \( z \), for \( n \to \infty \), uniformly in its components. In the next section we then use this successive approximation procedure to construct a better bounding function.

**Lemma 4:** Let \( z \) be defined as in lemma 1 and let (17) and (18) hold. Then, for all \( \epsilon > 0 \), there is a \( n_0 \) such that

\[ \| z - z_n \| < \epsilon, \quad \text{for} \ n > n_0 \]

**Proof:** Choose \( \epsilon > 0 \), \( \delta = \frac{\epsilon}{2 \| z \|} \)

As in the proof of lemma 2, we have, \( \forall a \in S \) a policy \( f_a \) such that

\[ z(a) \geq \sum_{n=0}^{\infty} \lambda^n \sum_{b \in S} (P^n(f))_{a,b} \geq z(a) - \delta. \] Hence

\[ 1 + \lambda \sum_{j \in S} p^a_j \left\{ \sum_{n=0}^{\infty} \lambda^n \sum_{b \in S} (P^n(f))_{j,b} \right\} \geq z(a) - \delta \]

and so:

\[ 1 + \lambda \sum_{j \in S} p^a_j z(j) \geq z(a) - \delta \]

Defining \( f \) as the policy with \( f(a) = f_a(a) \), \( a \in S \), we have

\[ z - \delta e \leq e + \lambda P(f) z \leq z \] (see also lemma 2) which implies

\[ e \geq z - \lambda P(f) z - \delta e \]

and so

\[ \sum_{n=0}^{N} \lambda^n P^n(f) e \geq z - \lambda^{N+1} P^{N+1}(f) z - \delta \sum_{n=0}^{N} \lambda^n P^n(f) e \]

while

\[ z = \sup_{\xi} \sum_{s}^{\infty} \lambda^n P_{r}^{n-1} P_{s} P(f) e \geq z \geq \sum_{n=0}^{N} \lambda^n P^n(f) e \]

So we have:

\[ z \geq z_N \geq z - \lambda^{N+1} P^{N+1}(f) z - \delta z. \]

Let \( N \) be so large that:
\[
\lambda^{N+1} P^{N+1}(f) z \preceq \lambda^{N+1} P^{N+1}(f) \|z\|e < \frac{\varepsilon}{2}
\]

Then
\[
z N + 1 = z N + \frac{1}{2} \varepsilon . e - \delta \|z\|e = z - \varepsilon . e
\]
or:
\[
\|z - z N\| < \varepsilon
\]
qu.e.d.

Remark: Because of \(z \succeq z_{N+1} \succeq z_N\), we have
\[
\|z - z_{N+1}\| < \varepsilon \Rightarrow \|z_{N+1} - z_N\| < \varepsilon
\]

3. Construction of a new bounding function

We just proved that the successive approximation procedure, described after lemma 3, converges uniformly to \(z\) under the conditions of lemma 1. Because of the remark at the end of section 2 we will find that, for \(N\) sufficiently large
\[
\|z_{N+1} - z_N\| < 1 \text{ or } z_{N+1} \preceq z_N + \varepsilon
\]
Because of \(z_{N+1} = \sup_{f} \{e + \lambda P(f)z_N\}\) we then have \(e + \lambda P(f)z_N \preceq z_N + \varepsilon\), for all \(f\), so
\[
(20) \ P(f)z_N \preceq \lambda^{-1}z_N, \text{ for all } f, \text{ or } \sum_{j \in S} P_{ij} z_N(j) \preceq \lambda^{-1}z_N(i), \forall i \in S, \forall k \in K
\]
If we write (20) in the notations of our original problem (that is: return to our "old" \(P(f)\) an \(\lambda\), and forget the transformation), we find
\[
M^{-1} \mu^{-1}(i) \sum_{j \in S} P_{ij} \mu(j) z_N(j) \preceq M^{-1} \lambda^{-1} z_N(i), \forall i \in S, \forall k \in K
\]
or: (21)
\[
\sum_{j \in S} P_{ij} \mu(j) z_N(j) \preceq \lambda^{-1} \mu(i) z_N(i) \quad \forall i \in S, \forall k \in K
\]
Defining \(\mu \circ z_N\) as the vector with elements \(\mu(i) z_N(i)\) we may write (21) as
\[
(22) \ P(f) (\mu \circ z_N) \preceq \lambda^{-1} (\mu \circ z_N), \quad \forall f
\]
Combining all these elements together we have proved the following theorem:

**Theorem:**

If for a Markov decision process there is a bounding function $\mu$ such that

\begin{equation}
\rho^* = \sup_{f} \limsup_{n \to \infty} \frac{1}{n} P^n(f) < 1
\end{equation}

(11)

\begin{equation}
M = \max\{1, \sup_{f} \|P(f)\| \mu\} < \infty
\end{equation}

(12)

then there is for each $\epsilon > 0$ a bounding function $\tilde{\mu}$ such that

\begin{equation}
P(f)\tilde{\mu} \leq (\rho^* + \epsilon)\tilde{\mu}
\end{equation}

(23)

\begin{equation}
\mu \leq \tilde{\mu} \leq L\mu \text{ for some constant } L
\end{equation}

(24)

Proof: Choose $\lambda = (\rho^* + \epsilon)^{-1}$, then the conditions of lemma 1 until 4 (after transforming the process) are fullfilled, so that we may apply the successive approximation procedure, described in §2, and take $\tilde{\mu} = \mu \circ z_N$, where $\mu \circ z_N$ satisfies (22).

4. Comments

Of course in practice one will not always know the spectral radius of a Markov decision process. In that case we must have at our disposal a good estimation of $\rho^*$, or a procedure, which results in a good bounding function, without knowing exactly the spectral radius. One of the things we can do is the following:

Suppose we have a Markov decision problem and a bounding function $\mu$ such that (12) holds. Furthermore let $\rho^*$, defined in (10) be smaller than one, but otherwise unknown.

Take $\lambda = 1$

Start the procedure, described in section 2, and continue until

\begin{equation}
\|z_{N+1} - z_N\| < \delta \times 1
\end{equation}

Then we have

\begin{equation}
e + \lambda P(f)z_N \leq z_{N+1} \leq z_N + \delta e
\end{equation}
or:

\[ \lambda P(z_N) \leq z_N - \delta \epsilon \]

and, because of \( \|z\| \leq \|z\| < \infty \)

\[ \lambda P(z_N) \leq z_N (1 - \frac{1 - \delta}{\|z_N\|}) = \gamma z_N' \quad \text{with} \quad \gamma = 1 - \frac{1 - \delta}{\|z_N\|} < 1 \]

This implies \( \lambda \rho^* \leq \gamma < 1 \), so we can improve \( \lambda \), i.e. take \( \lambda' = \lambda \gamma^{-1} \), and apply the same procedure again.

As long as \( \lambda < \rho^{*-1} \), the procedure is converging, and we will always find a \( \gamma < 1 \), so that \( \lambda \) can be improved. It can be proved that this sequence of \( \lambda \)-values is converging to \( \rho^{*-1} \).

References


