Memorandum COSOR 80-08

Geometric convergence in average reward Markov decision processes

by

W.H.M. Zijm

Eindhoven, June 1980

The Netherlands
0. Abstract

Recently, Federgruen and Schweitzer [3] proved that in undiscounted Markov decision problems the value iteration method for finding maximal gain policies converges geometrically fast, whenever convergence occurs. This result was obtained without any restriction on either the periodicity or chain structure of the problem. In this paper we establish the same result once again; the proof however, seems essentially simpler and, moreover, yields an upperbound for the convergence rate.

1. Introduction

In a recent, remarkable paper [3] Federgruen and Schweitzer showed that the value iteration method for finding maximal gain policies in undiscounted Markov decision problems exhibits a geometric rate of convergence, whenever convergence occurs. In the case that after a finite number of steps we are dealing with only one maximizing policy, this fact is immediately clear by exploiting the so-called Jordan-form of a matrix (compare e.g. Pease [5]). In this case a sharp upperbound for the ultimate convergence factor is given by the absolute value of the largest eigenvalue with radius strictly smaller than one (which is not the same as the subradius as it was defined by e.g. Morton and Wecker [4]; in multichain or periodic cases this subradius is equal to one). In this paper we present an alternative proof for the main result in Federgruen and Schweitzer [3]; a proof however, which we believe to be essentially simpler. Moreover, we find an upper bound for the ultimate convergence rate.
We consider a discrete time Markov decision process with finite state space \( S = \{1, \ldots, N\} \) and finite action space \( K \). Choosing action \( k \in K \) when the system is in state \( i \in S \) results in a probability \( p_{ij}^k \) of observing the system in state \( j \) at the next point of time. Furthermore a reward \( r_{ij} \) is earned. A policy \( f \) is a function from the state space to the action space, a strategy \( \pi \) is a sequence of policies: \( \pi = (f_0, f_1, f_2, \ldots) \). A strategy \( \pi = (f, f, f, \ldots) \) is called stationary. Furthermore we denote by \( P(f) \) the matrix with elements \( P_{ij}^k \); \( i, j = 1, \ldots, N; \) \( r(f) \) denotes the vector with components \( r_{i}^f ; i = 1, \ldots, N \).

In undiscounted Markovian decision problems the value-iteration equations can be written as follows:

\[
(1) \quad v_i(n) = \max_{k \in K} \left\{ r_i^k + \sum_{j=1}^{N} p_{ij}^k v_j(n-1) \right\}; \quad i = 1, \ldots, N; \quad n = 1, 2, \ldots.
\]

Here \( v_i(n) \) denotes the \( i \)-th component of a vector \( v(n) \) and \( v(0) \) is given. Brown [1] showed that

\[
(2) \quad \| v(n) - ng^* \| \leq C
\]

for some constant \( C \) (here \( \| \ldots \| \) denotes the usual sup-norm). In the above expression \( g^* \) denotes the maximal gain rate vector, defined by

\[
(3) \quad g_i^* = \max_f g_i(f) ; \quad i = 1, \ldots, N
\]

with

\[
g(f) = \lim_{n \to \infty} \frac{1}{n+1} \sum_{\ell=0}^{n} P_{\ell}^f r(f).
\]

Derman [2] proved that there exists a policy that achieves the \( N \) suprema in (3) simultaneously.

In general \( \{v(n) - ng^*_i\}_{n=1}^{\infty} \) may fail to converge for arbitrary \( v(0) \) if some of the transition probability matrices are periodic. The necessary and sufficient condition for the convergence of \( \{v(n) - ng^*_i\}_{n=1}^{\infty} \) for all \( v(0) \) was obtained by Schweitzer and Federgruen [6]. A very easy sufficient condition is the
assumption that all matrices are aperiodic. The main result in Federgruen and Schweitzer [3] finally states that, whenever \( \lim_{n \to \infty} \{ v(n) - n g^* \} \) exists, the convergence to the limit is geometric. It is this result that will be proved here again in a relatively simple way, in section 2. The method of proving the result implies the existence of an upper-bound, strictly smaller than one, for the ultimate convergence rate, which is independent of the starting vector.

2. Geometric convergence of value-iteration

As a starting point in this section we assume that \( \{ v(n) - n g^* \} \) is converging to some vector \( w^* \). Define \( \{ e(n) \} \) by

\[
e(n) = v(n) - n g^* - w^* , \quad n = 1, 2, \ldots .
\]

Our aim in this section will be to prove that \( e(n) \) is approaching zero geometrically fast after a finite number of steps, i.e. for \( n \geq n_0 \) say. Substitution of (4) into (1), and writing (1) as a vector-equation gives, for \( n = 1, 2, \ldots \)

\[
ge^* + w^* + e(n) = \max_f \{ r(f) + (n-1) P(f) g^* + P(f) w^* + P(f) e(n-1) \}.
\]

Divide both sides of (5) by \( n \) and let \( n \) tend to infinity. Then we find

\[
\max_f P(f) g^* = g^* .
\]

Let \( A \) denote the set of policies which maximize \( P(f) g^* \). It is clear that for \( n \) sufficiently large, \( n \geq n_1 \) say, a maximizing policy \( f \) in (5) will be in \( A \), hence \( P(f) g^* = g^* \). Subtracting this from (5) reduces the functional equation to

\[
g^* + w^* + e(n) = \max_{f \in A} \{ r(f) + P(f) w^* + P(f) e(n-1) \} ; \quad n \geq n_1
\]

Again, let \( n \) tend to infinity. Then we have, since \( \lim_{n \to \infty} e(n) = 0, \)

\[
g^* + w^* = \max_{f \in A} \{ r(f) + P(f) w^* \} .
\]
Let $B$ denote the set of policies which maximize $r(f) + P(f)w^*$ (hence $B \subseteq A$).

For $n$ sufficiently large, $n \geq n_2$ say, we find with the same arguments as above the following reduced functional equation

$$e(n) = \max_{f \in B} P(f)e(n-1) ; \quad n \geq n_2$$

(9)

(obviously $n_2 \geq n_1$).

It will be our problem to prove that $\lim_{n \to \infty} e(n) = 0$ implies that the convergence to zero is also geometrically fast. For generality we reformulate the problem as follows:

Suppose we have a set of (column) vectors \( \{x(n); n = 0, 1, \ldots \} \) which obey the following dynamic programming recursion

$$x_i(n) = \max_{k \in K} \sum_{j=1}^{N} p_{ij}^k x_j(n-1) ; \quad n = 0, 1, 2, \ldots ; \quad i = 1, \ldots, N.$$  \hspace{1cm} (10)

where $K$ is a finite set of actions, as defined in the introduction. Suppose furthermore

$$\lim_{n \to \infty} x(n) = 0 .$$  \hspace{1cm} (11)

Then we want to establish the geometric convergence of \( \{x(n); n = 0, 1, \ldots \} \).

We will need the following definitions:

$$C_n = \{i \mid i \in S ; \quad x_i(n) > 0\} ; \quad C'_n = S \setminus C_n ,$$

$$D_n = \{i \mid i \in S ; \quad x_i(n) < 0\} ; \quad D'_n = S \setminus D_n .$$

For a strategy $\pi = (f_1, f_2, f_3, \ldots)$ we define $P^{(n)}_\pi$ as

$$P^{(n)}_\pi = P(f_n) P(f_{n-1}) \ldots P(f_1) .$$
From (10) and (11) it is clear that \( C'_n \neq \emptyset, D'_n \neq \emptyset \), for all \( n \). Furthermore \( C_n \subseteq D' \), \( D_n \subseteq C' \), for all \( n \). The following lemma asserts that \( \max \{ x_i(n) \} \) is decreasing to zero geometrically fast, whenever \( C_0 \neq \emptyset \).

**Lemma 1**: Suppose \( C_0 \neq \emptyset \) and let (10) and (11) hold. Then there exist numbers \( \alpha \in \mathbb{R}, 0 \leq \alpha < 1 \) and \( n_0 \in \mathbb{N}, n_0 \leq 2^N \) such that

\[
\max_{i \in S} \{ x_i(n) \} \leq \alpha^k \max_{i \in S} \{ x_i(0) \}.
\]

**Proof**: Define \( R_0 = C_0 \) and \( R_n \) by

\[
R_n = \{ i \mid \exists k \in K: \sum_{j \in R_{n-1}} p_{ij}^k = 1 \} ; \quad n = 1, 2, \ldots.
\]

Then it follows immediately that \( R_n \subseteq C \). Now suppose that \( R_n \neq \emptyset \) for some \( n \geq 2^N \). Since there are at most \( 2^N - 1 \) non-empty subsets of \( S \) we must have for some \( n_1, n_2 \in \mathbb{N} \) with \( 0 \leq n_1 < n_2 \leq 2^N \) that \( R_{n_1} = R_{n_2} \). Define

\[
R = R_{n_1} = R_{n_2}.
\]

By definition of \( R_n \) there exists a finite sequence of policies \( \{ f_{n_1}, \ldots, f_{n_2} \} \) such that

\[
\sum_{j \in R} [p(f_{n_1}) \ldots p(f_{n_2})]_{ij} = 1 \text{ for } i \in R.
\]

Let \( \bar{x}(n_1) = \min_{i} x_i(n_1) \). By repeating the sequence \( \{ f_{n_1}, \ldots, f_{n_2} \} \) again and again we immediately conclude

\[
x_i(n_1 + k(n_2 - n_1)) \geq \bar{x}(n_1) > 0 \quad k = 1, 2, \ldots ; \quad i \in R
\]

which contradicts (11). Hence \( R_n = \emptyset \) for \( n \geq 2^N \).
Now define $n'_0 \in \mathbb{N}$ and $a' \in \mathbb{R}$ as follows:

\[
n'_0 := \min \{ n \mid R_n = \emptyset \}
\]

\[
a' := \max \max \{ \sum_{j \in C_0} [P_{ij}]^{(n'_0)} \mid i \in S \}
\]

then $a' < 1$ and we find

\[
\max \{ x_i(n'_0) \} \leq a' \max \{ x_i(0) \} .
\]

We may reason in the same way, using $C_1$ as a starting point; we then find numbers $n''_0$, $a''$, etc., etc. Since $S$ possesses at most $2^N - 1$ non-empty subsets it follows that we are able to determine numbers $n_0 \in \mathbb{N}$, $n_0 \leq 2^N$ and $a \in \mathbb{R}$, $a < 1$, such that

\[
\max \{ x_i(n + n_0) \} \leq a \max \{ x_i(n) \} ; \quad n = 0, 1, 2, ...
\]

and hence

\[
\max \{ x_i(k \cdot n_0) \} \leq a^k \max \{ x_i(0) \} ; \quad k = 1, 2, ...
\]

An analogous result may be found for $\min \{ x_i(n) \}$. We have

\[
\min \{ x_i(n) \} = \min \{ x_i(n_0) \} \leq a^k \min \{ x_i(0) \} .
\]

Lemma 2 : Suppose $D_0 \neq \emptyset$ and let (10) and (11) hold. Then there exist numbers $\beta \in \mathbb{R}$, $0 \leq \beta < 1$ and $m_0 \in \mathbb{N}$, $m_0 \leq 2^N$ such that

\[
\min \{ x_i(k \cdot m_0) \} \geq \beta^k \min \{ x_i(0) \} .
\]

Proof : Define $T_0 = D_0$ and $T_n$ by

\[
T_n = \{ i \mid \sum_{j \in T_{n-1}} P_{ij}^k = 1 \text{ for all } k \in K \} ; \quad n = 1, 2, ...
\]
Then obviously \( T_n \subset D_n \) and \( T_n = \emptyset \) for \( n \geq 2^N \), since the opposite assertion would imply, by an argument analogous to the one in the proof of Lemma 1, the existence of integers \( n_1, n_2 \) with \( 0 \leq n_1 < n_2 \leq 2^N \) such that 
\[
T_{n_1} = T_{n_2} =: T \text{ and by definition of } T_n :
\]
\[
\sum_{j \in T} [P_{\pi}]_{ij} = 1 \text{ for } i \in T, \text{ for all } \pi.
\]
Then with \( \hat{x}(n_1) = \max_{i \in T} x_i(n_1) \) we would find
\[
x_i(n_1 + k(n_2 - n_1)) \leq \hat{x}(n_1) < 0 \quad ; \quad k = 1, 2, \ldots ; \quad i \in T,
\]
contradicting (11) again.
Hence we may define \( m'_0 \in \mathbb{N} \) and \( \beta' \in \mathbb{R} \) as follows
\[
m'_0 := \min \{ n \mid T_n = \emptyset \} ,
\]
\[
\beta' := \max \min_{i \in S} \{ \sum_{j \in D_0} (m'_0)_{ij} \} ;
\]
then \( \beta' < 1 \) and we find
\[
\min_{i} \{ x_i(m'_0) \} \geq \beta' \cdot \min_{i} \{ x_i(0) \} .
\]
Using an analogous argument as in the foregoing proof, we also may determine numbers \( m_0 \in \mathbb{N} , \quad m_0 \leq 2^N \) and \( \beta \in \mathbb{R} , \quad \beta < 1 \) such that
\[
\min_{i} \{ x_i(n + m_0) \} \geq \beta \cdot \min_{i} \{ x_i(n) \} ; \quad n = 0, 1, 2, \ldots
\]
and hence
\[
\min_{i} \{ x_i(k \cdot m_0) \} \geq \beta^k \cdot \min_{i} \{ x_i(0) \} ; \quad k = 1, 2, \ldots
\]
Lemma 1 and Lemma 2 together establish the geometric convergence of 
\{x(n) ; n = 1,2,...\} to zero. Notice that \(a\) and \(B\) do not depend explicitly 
on the vectors \(x(n), n = 1,2,...\), but merely on the positions of \(C_n\) 
and \(D_n\), and on the probabilistic behaviour of the set of matrices. Since 
\(S\) contains only a finite number of the non-empty subsets and moreover, we 
have only a finite number of policies (and hence a finite number of strategies 
of length \(\leq 2^N\)) this implies that \(n_0, m_0\), \(a\) and \(B\) may be chosen indepen-
dent of \(x(0)\), that is: the geometric convergence is uniform with re-
spect to the starting vector, while \(\max(a,B)\) constitutes an upperbound for 
the convergence rate.

We further remark that the finiteness of \(K\) is not very essential; under the 
following condition we still have geometric convergence with a factor, 
strictly smaller than one:

\[
\min_{i,j,k} \{p_{ij}^k \mid p_{ij} > 0\} = \epsilon > 0 .
\]
References


