Successive approximations for convergent dynamic programming
van Hee, K.M.; Hordijk, A.; van der Wal, J.

Published: 01/01/1977

Citation for published version (APA):
Successive Approximations for
Convergent Dynamic Programming

by
Kees M. van Hee, Arie Hordijk and
Jan van der Wal

Eindhoven, April 1977
The Netherlands
Successive Approximations for Convergent Dynamic Programming

by

Kees M. van Hee, Arie Hordijk and Jan van der Wal

1. Introduction and Preliminaries

The main topic of this paper is the convergence of the method of successive approximations for dynamic programming with the expected total return criterion.

We first sketch the framework of the dynamic programming model we are dealing with.

Consider a countable set \( E \), the state space, and an arbitrary set \( A \), the action space, endowed with some \( \sigma \)-field containing all one-point sets. Let \( p \) be a transition probability from \( E \times A \) to \( E \) (notation: \( p(j|i,a) \), \( i,j \in E \), \( a \in A \)). Let \( H_n := (E \times A)^n \times E \) be the set of histories until time \( n \) (\( n \geq 1 \)) and \( H_0 := E \).

In all generality a strategy \( \pi \) is a sequence \((\pi_0, \pi_1, \ldots)\) where \( \pi_n \) is a transition probability from \( H_n \) to \( A \). The set of all strategies is denoted by \( \Pi \). The subset \( M \) of \( \Pi \) consists of all Markov strategies; i.e. \( \pi = (\pi_0, \pi_1, \ldots) \in M \) if and only if there is a sequence of functions \( f_0, f_1, \ldots, f_n : E + A, n = 0,1, \ldots \), such that

\[
\pi_0(f_0(i)|i) = 1, \quad \pi_n(f_n(i)|h_{n-1}, a_{n-1}, i) = 1
\]

for all \( h_{n-1} \in H_{n-1}, a_{n-1} \in A, \) and \( i \in E \). Each \( i \in E \) and \( \pi \in \Pi \) determine a probability \( \mathbb{P}_{i,\pi} \) on \((E \times A)^\infty\) and a stochastic process \( \{(X_n, Y_n), n = 0,1, \ldots\} \) where \( X_n \) is the state and \( Y_n \) the action at time \( n \). The expectation with respect to \( \mathbb{P}_{i,\pi} \) is denoted by \( \mathbb{E}_{i,\pi} \).

The reward function \( r \) is a real measurable function on \( E \times A \).

Throughout this paper we assume

\[
\sup_{\pi \in M, i} \mathbb{E}_{i,\pi} \left[ \sum_{n=0}^{\infty} r^+(X_n, Y_n) \right] < \infty \quad \text{for all } i \in E
\]
(note that $x^+ := \max(x, 0)$). This assumption guarantees that the expected total return $v(i, \pi) := \mathbb{E}_{i, \pi} \left[ \sum_{n=0}^{\infty} r(X_n, Y_n) \right]$ is defined for all $i \in E$ and $\pi \in \Pi$, and in [9] it is proved, using a well-known theorem of Derman and Strauch [4], that

$$\text{(1.2)} \quad \sup_{\pi \in \Pi} v(i, \pi) = \sup_{\pi \in \Pi} v(i, \pi) \quad \text{for all } i \in E.$$ 

As a consequence of 1.2 we are mainly interested in Markov strategies and for that reason we introduce some notations which are especially useful for this class.

First we define the set $\mathcal{P}$ of transition probabilities from $E$ to $E$: for which there is a function $f : S \rightarrow A$ such that $P(i, \cdot) = p(\cdot | i, f(i))$. For all $i \in S$ and further a function $r : E \times \mathcal{P} \rightarrow \mathbb{R}$ (= the set of reals)

$$r_P(i) := \sup \{r(i, a) | P(i, \cdot) = p(i, a, \cdot), a \in A \}.$$ 

Note that each $\pi \in \Pi$ is completely determined by a sequence $R = (P_0, P_1, \ldots)$, $P_n \in \mathcal{P}$, $n = 0, 1, \ldots$. Hence we may identify each $\pi \in \Pi$ with such a sequence $R$, and express

$$E_{i, R} r(X_n, Y_n) = P_0 \cdots P_{n-1} r_P(i), \quad \text{for } R = (P_0, P_1, \ldots), \quad i \in E.$$ 

(By convention the empty product of elements of $\mathcal{P}$ is the identity operator, and if we omit the subscript $i$ in $E_{i, R}$ we mean the function on $E$).

On $E$ we define the functions:

$$\text{(1.3)} \quad v := \sup_{R \in \Pi} E_R \left[ \sum_{n=0}^{\infty} r(X_n, Y_n) \right], \quad \text{the value function}$$

for a function $s : E \rightarrow \mathbb{R}$ with $\sup_{R \in \Pi} E_R [s^+(X_k)] < \infty$, $k = 0, 1, \ldots$.

$$\text{(1.4)} \quad v^s := \sup_{R \in \Pi} E_R \left[ \sum_{k=0}^{n-1} r(X_k, Y_k) + s(X_n) \right], \quad v_n := v^0_n$$

for a sequence $a = (a_0, a_1, \ldots)$ of functions $a_n : E \rightarrow \mathbb{R}$, $\hat{\mathbb{R}} := \{x \in \mathbb{R} | x \geq 1\}$ we define the function $w_a$ and $z_a$ on $E$:  

A dynamic programming model is said to be stable with respect to a scrap function \( s \) if:

\[
\lim_{n \to \infty} v_n^s(i) = v(i), \quad \text{for all } i \in E
\]

It is well known that positive, negative and discounted dynamic programming models with finite \( E \) and \( A \) are stable. But this is not true in convergent dynamic programming, the case that \( z \) is finite (see [13], [14]), as is shown by the following example.

Counterexample:

\[
E = \{1, 2\}, \quad A = \{1, 2\}, \quad p(1|1,1) = p(2|1,2) = 1, \quad r(1,1) = 0, \quad r(1,2) = 2, \\
p(\cdot|2,1) = p(\cdot|2,2) = 0, \quad r(2,1) = r(2,2) = -1.
\]

Then \( v_n^0(1) = 2 \) and \( v(1) = 1 \).

It is well-known that stability (with respect to scrap function 0) is guaranteed, if the expected total return from time \( n \) onwards, tends to zero as \( n \) tends to infinity uniformly in the strategy. In 1.8 this uniform tail property is defined:
In this paper two types of assumptions are considered to guarantee this uniform tail convergence. In section 2 the strong convergence conditions are introduced. A model is called strongly convergent if $w_a$ or $z$ is finite for a sequence of functions $a = (a_0, a_1, \ldots)$ with $\lim_{n \to \infty} a_n(i) = \infty$ for all $i \in E$. It turns out that property 1.8 is equivalent to a strong convergence condition. In section 3 Liapunov functions are introduced and the existence of finite Liapunov functions is related to strong convergence. In section 4 Liapunov functions turn out to be important tools in successive approximations because they provide bounds for $|v - v^S_n|$ and procedures for excluding suboptimal actions.

In section 5 the connection with contracting dynamic programming is made and in section 6 a waiting line model with controllable input is presented, which satisfies the strong convergence condition but which is not contracting. Finally in section 7 some results on (nearly) optimal strategies are collected.

We conclude this section with some remarks and notations. Models with for each $i \in E$ a different action space $A_i$ can easily be transformed into our framework.

In [13] and [14] convergent dynamic programming ($z < \infty$) was studied extensively. In this paper we are almost always working within this framework, since besides the overall assumption 1.1 we work with additional assumptions which are at least as strong as: $w$ is finite. Hence with $w < \infty$ and

$$\limsup_{n \to \infty} \sum_{n=0}^{\infty} |E_{i,R}[r(X_n, Y_n)]| = 0$$

we have

$$z(i) \leq 2 \sup_{R \in M} \sum_{n=0}^{\infty} r^+(X_n, Y_n) + w(i) < \infty.$$
a ≤ x iff a(i) ≤ x for all i ∈ E (the same holds if ≤ is replaced by < or =). With the convergence of a sequence of functions on E we mean pointwise convergence and the supremum of a collection of functions is the pointwise supremum. With convergence of a sequence of elements of P we mean elementwise convergence. For an extended real valued function a and a positive function b on E we write \( \frac{a}{b} \) for the function \( c(i) := \frac{a(i)}{b(i)} \).

For a nonnegative function \( \mu \) on E we introduce the set

\[ V(\mu) := \{ v ∈ \mathbb{R}^∞, |v| ≤ k\mu \text{ for some } k ∈ \mathbb{R} \}. \]

On \( V(\mu) \) we define that the norm \( \mu \) by

\[ \|f\|_\mu = \sup\{\mu^{-1}(i)|f(i)|, i ∈ E, \mu(i) > 0\}. \]

The function \( \mu \) is called a bounding function (c.f. section 5). For functions \( f \) on E with

\[ \sup\{ Pf \} < ∞ \quad P ∈ P \]

we define two wellknown operators

\[ U_f := \sup\{ r + Pf \} \quad P ∈ P \]

\[ (1.10) \]

\[ \hat{U}_f := \sup Pf \quad P ∈ P \]

Finally we formulate Bellman's optimality equations:

\[ (1.11) \quad v^s_n = U^n_s \]

\[ (1.12) \quad v = Uv \]

The Liapunov-approach was presented by Hordijk at the Advanced Seminar on Markov Decision Theory, Amsterdam 1976. So he inspired van Hee and van der Wal to investigate the problem of successive approximations under very general conditions, which resulted in the strong convergence-approach. Then the three of us joined the investigations which led to this paper.
2. Strong convergence

One of the main results in this section is the equivalence of the strong convergence condition with the uniform tail property expressed in 1.8.

We first give some simple, but useful inequalities.

Throughout this section let \( a = (a_0, a_1, \ldots) \) be a nondecreasing sequence of functions, \( a_n : \mathbb{E} \to \mathbb{R} \).

**Theorem 2.1.**

(i) \[ \sup_{R \in \mathcal{M}} \sum_{k=n}^{\infty} |\mathbb{E}_{R} r(X_k, Y_k)| \leq \frac{w}{a_n} \]

(ii) \[ \sup_{R \in \mathcal{M}} \sum_{k=n}^{\infty} \mathbb{E}_{R} |r(X_k, Y_k)| \leq \frac{z}{a_n} \]

**Proof.** Since \( a_k(i) \) is nondecreasing in \( k \) and \( a_1(i) > 0 \), we have, for all \( i \in \mathbb{E} \):

\[ \sup_{R \in \mathcal{M}} \sum_{k=n}^{\infty} |\mathbb{E}_{i, R} r(X_k, Y_k)| \leq \sup_{R \in \mathcal{M}} \sum_{k=n}^{\infty} \frac{a_k(i)}{a_n(i)} |\mathbb{E}_{i, R} r(X_k, A_k)| \leq \frac{w(i)}{a_n(i)} . \]

The proof of (ii) is similar.

**Lemma 2.2.**

(2.1) \[ \sup_{R \in \mathcal{M}} \mathbb{E}_{R} z(X_n) = \sup_{R \in \mathcal{M}} \mathbb{E}_{R} \left[ \sum_{k=n}^{\infty} |r(X_k, Y_k)| \right] = \overline{U}^n z \]

**Proof.**

\[ \sup_{R \in \mathcal{M}} \mathbb{E}_{R} z(X_n) = \sup_{R \in \mathcal{M}} P_0 \cdots P_{n-1} z = \overline{U}^n z . \]

And further:

\[ \sup_{P_0 \cdots P_{n-1} z} = \sup_{P_0 \cdots P_{n-1}} \left| \sum_{k=0}^{\infty} P_{n+k} \right| = \sup_{P_0 \cdots P_{n-1}} \left| \sum_{k=0}^{\infty} P_{n+k} \right| = \sup_{P_0, P_1, \ldots} \left| \sum_{k=0}^{\infty} r(X_k, Y_k) \right| . \]
A direct consequence of theorem 2.1 and lemma 2.2 is

\[ (2.2) \quad \sup_{R \in M} E_R |v(X_n)| \leq \sup_{R \in M} E_R z(X_n) \leq \frac{z}{a_n}. \]

And in a similar way one may prove

\[ (2.3) \quad \sup_{R \in M} |E_R v(X_n)| \leq \frac{w_a}{a_n}. \]

One of the consequences of the above inequalities is that if \( z_a < \infty \) for some sequence \( a \) with

\[ \lim_{n \to \infty} a_n = \infty \]

then

\[ \lim_{n \to \infty} E_R |v(X_n)| = 0 \]

for any strategy. Hence any strategy is equalizing (see chapter 4 of [13]). See also theorem 7.8.

Theorem 2.3. states that \( w_a < \infty \) and \( \lim_{n \to \infty} a_n = \infty \) guarantee stability. Note that \( w_a \leq z_a \).

Theorem 2.3.

Let \( w(a) < \infty \) and \( \lim_{n \to \infty} a_n = \infty \). Then the problem is stable with respect to any scrapfunction \( s \) satisfying \( \sup_{R \in M} E_R s(X_n) < \infty, \ n = 0, 1, \ldots \) and \( \sup_{R \in M} E_R s(X_n) \to 0 \) \( (n \to \infty) \).

Proof.

\[ v - v_n^s = \sup_{R \in M} E_R \left[ \sum_{k=0}^{n-1} r(X_k, Y_k) \right] + \sup_{R \in M} E_R \left[ \sum_{k=0}^{n} r(X_k, Y_k) + s(X_n) \right] \]

\[ \leq \sup_{R \in M} E_R \left[ \sum_{k=n}^{\infty} r(X_k, Y_k) \right] + \sup_{R \in M} E_R s(X_n) \]

\[ \leq \frac{w}{a_n} + \sup_{R \in M} E_R s(X_n). \]
Similarly one shows
\[ \nu^s_n - \nu \leq \frac{w_n}{a_n} + \sup_{R \in M} \left| \mathbb{E}_R s(X_n) \right| . \]
Hence \( \lim_{n \to \infty} |\nu^s_n - \nu| = 0 . \)

So theorem 2.3 gives a new criterion for stability.

If \( z_a < \infty \) and \( \lim_{n \to \infty} a_n = 0 \) we may use scrapfunctions \( s \) satisfying for some \( K \in \mathbb{R} \)
\[ |s| \leq Kz , \]
since by theorem 2.1 and lemma 2.2 \( \limsup_{n \to \infty} \mathbb{E}_R z(X_n) = 0 . \)

Consider a dynamic programming model with bounded rewards, say \( |r(i,a)| \leq b \) for all \( i \in E, a \in A \) and let \( E_0 \) be an absorbing subset of \( E \) with \( r(i,a) = 0 \) for all \( i \in E_0, a \in A \). Let \( T \) be the entrance time in \( E_0 \). If \( \sup_{R \in M} \mathbb{E}_R T < \infty \) then this model satisfies the strong convergence condition in a natural way since
\[ z_a \leq b \sup_{R \in M} \mathbb{E}_R T \text{ for } a_n \equiv n + 1, \ n = 0, 1, \ldots . \]

In fact \( |\nu_n - \nu| \leq \frac{1}{n+1} b \sup_{R \in M} \mathbb{E}_R T \). Similar expressions can be derived with higher moments of the entrance time.

In general one may say if \( w(a) < \infty \) then \( |\nu_n(i) - \nu(i)| \) tends to zero at a rate at least as fast as \( [a_n(i)]^{-1} \).

From the foregoing results the question arises under which conditions there exists a sequence of functions \( a \) with \( a_n \to \infty \) and \( w \to \infty \). The following theorem gives the already announced characterization.

**Theorem 2.4.**

There exists a nondecreasing sequence of functions \( a = (a_0, a_1, \ldots) \) on \( E \) with \( \lim_{n \to \infty} a_n = \infty \) and \( w \to \infty \) if and only if
\[ w < \infty \text{ and } \limsup_{n \to \infty} \sup_{R \in M} \sum_{k=n}^{\infty} \mathbb{E}_{i,R} |r(X_k, Y_k)| = 0 . \]
Proof. First the if part. Define

\[ b_n(i) := \sup_{\mathbb{R}^m} \sum_{k=n}^{\infty} |E_{i,k}[r(X_k, Y_k)]|, \quad i \in E. \]

Obviously, \( b_n \geq b_{n+1} \). Now let \( a_n(i) = \ell + 1 \) if \( N_{\ell}(i) \leq n < N_{\ell+1}(i) \) with \( N_0(i) := 0 \) and \( N_{\ell}(i) := \min\{n \mid b_n(i) \leq 2^{-\ell}\}, \ell = 1, 2, \ldots \). Then

\[ N_{\ell+1}(i) \sup_{\mathbb{R}^m} \sum_{n=N_{\ell}(i)}^{\infty} a_n(i) |E_{i,n}[r(X_n, Y_n)]| \leq (\ell + 1)2^{-\ell}, \ell = 1, 2, \ldots \]

and consequently

\[ \omega_a(i) \leq \sup_{\mathbb{R}^m} \sum_{n=0}^{\infty} |E_{i,n}[r(X_n, Y_n)]| + \sum_{\ell=1}^{\infty} (\ell + 1)2^{-\ell} \leq \omega + 3 < \infty. \]

The only if part is immediate from \( \omega \leq \omega(a) < \infty \) and theorem 2.1.(i). □

In theorem 2.5 we collect two sufficient conditions for stability which are weaker than the strong convergence condition. It is well known that positive dynamic programming models are stable, but the strong convergence condition need not be fulfilled there. The following theorem covers also the positive case.

Theorem 2.5.

Each of the following conditions guarantees stability for scrapfunction 0.

(i) \( \liminf_{n \to \infty} \inf_{\mathbb{R}^m} E_{X_n} v(X_n) \geq 0 \)

(ii) there exists a nondecreasing sequence \( a = (a_0, a_1, \ldots) \) of functions and

\[ a_n : E \to \mathbb{R} \text{ with } \lim_{n \to \infty} a_n = \infty \]

\[ d_a(i) := \sup_{\mathbb{R}^m} E_{i,n}[\sum_{n=0}^{\infty} a(i)r(X_n, Y_n)] < \infty \quad (x^- = \max(0, -x)). \]
Proof. For all $R \in M$

$$v_n \geq \mathbb{E}_R \left[ \sum_{k=0}^{n-1} r(X_k, Y_k) \right].$$

Hence

$$\liminf_{n \to \infty} v_n \geq v(\cdot, R) \quad \text{for all } R \in M$$

and consequently

$$\liminf_{n \to \infty} v_n \geq \sup_{R \in M} v(\cdot, R) = v.$$

Hence to prove stability we have to show

$$\limsup_{n \to \infty} v_n \leq v.$$

Part (i). By the optimality equation we have $r_p + Pv \leq v$, $P \in P$.

Hence by iteration

$$\sum_{k=0}^{n-1} P_0 \ldots P_{k-1} r_{P_k} + P_0 \ldots P_{n-1} v \leq v$$

or

$$\mathbb{E}_R \left[ \sum_{k=0}^{n-1} r(X_k, Y_k) \right] + \mathbb{E}_R [v(X_n)] \leq v$$

Consequently,

$$v_n + \inf_{R} \mathbb{E}_R [v(X_n)] \leq v$$

So with

$$\liminf_{n \to \infty} \inf_{R} \mathbb{E}_R [v(X_n)] \geq 0$$

we find

$$\limsup_{n \to \infty} v_n \leq v.$$
Part (ii). For $R \in M$

$$v(i,R) = \mathbb{E}_R[\sum_{k=0}^{n-1} r(X_k, Y_k)] + \mathbb{E}_R[\sum_{k=n}^{\infty} r(X_k, Y_k)] \geq \mathbb{E}_R[\sum_{k=0}^{n-1} r(X_k, Y_k)] - \sup_{R \in M} \mathbb{E}_R[\sum_{k=n}^{\infty} r^-(X_k, Y_k)]$$

Hence, by taking the supremum over $R \in M$

$$v \geq v_n - \sup_{R \in M} \mathbb{E}_R[\sum_{k=n}^{\infty} r^-(X_k, Y_k)].$$

Using 2.4 one proves in a way similar as in the proof of theorem 2.1

$$\sup_{R \in M} \mathbb{E}_R[\sum_{k=n}^{\infty} r^-(X_k, Y_k)] \leq \frac{d}{a_n}.$$  

Hence

$$v \geq \limsup_{n \to \infty} v_n - \lim_{n \to \infty} \frac{d}{a_n} = \limsup_{n \to \infty} v_n.$$  

If for some sequence $P_0, P_1, \ldots$ we have

$$v_n = r_{P_n} + P_n v_{n-1}$$

then

$$\liminf_{n \to \infty} P_n \ldots P_0 v \geq 0$$

is sufficient for stability, since iteration of the inequality $v \geq r_P + P v$ yields

$$v \geq r_{P_n} + \sum_{k=1}^{n} P_{n-k+1} r_{P_{n-k}} + P_{n-k} \ldots P_0 v = v_n + P_n \ldots P_0 v$$

and the proof of theorem 2.5 $\limsup_{n \to \infty} v_n \leq v$ is sufficient for stability.
3. Liapunov Functions and Strong Convergence

We first introduce Liapunov functions.
Consider a sequence of nonnegative extended real functions \( \ell_1, \ell_2, \ldots \) on \( E \) satisfying for all \( P \in \mathcal{P} \) the inequalities

\[
\ell_1 \geq |r_P| + P\ell_1
\]

(3.1)

\[
\ell_k \geq \ell_{k-1} + P\ell_k, \quad k = 2, 3, \ldots
\]

Finite solutions of (3.1) are called Liapunov functions. If \( \ell_k \) is finite, \( \ell_k \) is called a Liapunov function of order \( k \). Note that \( \ell_k < \infty \) implies \( \ell_{k-1} < \infty \).
Liapunov functions are powerful tools in dynamic programming. They were first studied in a context of dynamic programming in [13] chapter 4 for the convergent dynamic programming model and in chapter 5 of [13] and in [15] Liapunov functions are studied in connection with the average return criterion for models in which some state is recurrent under each strategy and in [14] they are used to obtain (partial) Laurent expansions for the expected total discounted return. In section 4 the existence of a Liapunov function of order 2 is assumed to obtain bounds for \( |v^n_s - v| \).

The functions \( y_1, y_2, \ldots \) defined in 1.7 satisfy Bellman's optimality equation, hence

\[
y_1 = \sup_{P \in \mathcal{P}} \{|r_P| + Py_1\}, \quad \text{if } y_1 < \infty
\]

and

\[
y_k = \sup_{P \in \mathcal{P}} \{y_{k-1} + Py_k\}, \quad \text{if } y_k < \infty, \quad k = 2, 3, \ldots
\]

Hence, if \( y_k \) is finite, \( y_1, \ldots, y_k \) are Liapunov functions and moreover it is easy to verify that \( \ell_k < \infty \) implies \( \ell_n \geq y_n \), \( n = 1, 2, \ldots, k \). Although we can work with \( y_k \) in stead of \( \ell_k \) for theoretical purposes it may happen in applications that one can find, in a relative simple way, Liapunov functions \( \ell_1, \ell_2, \ldots, \ell_k \), while the functions \( y_1, y_2, \ldots, y_k \) are hard to obtain. Since there is a large class of Liapunov functions there still is some freedom to choose an appropriate one. Specially this might improve the bounds in the approximation procedure (see also section 4). In this section we concentrate on the relations between Liapunov functions and strong convergence.
We recall that the finiteness of a Liapunov function of order $k$ is equivalent to the finiteness of $y_1, \ldots, y_k$.

**Theorem 3.1.**

$$y_n \geq \sup_{R \in M} \mathbb{E}_R \sum_{k=0}^{\infty} \binom{k+n-1}{k} |r(X_k, Y_k)|$$

**Remark.**

Hence $y_n < \infty$ implies $z_a < \infty$ for a sequence functions $a_k \equiv \binom{k+n-1}{k}$, and consequently the strong convergence condition holds.

**Proof.** By induction. For $n = 1$ the statement holds by definition 1.7. Suppose it holds for $n - 1$ ($n \geq 2$) then:

$$y_n = \sup_{P_0, \ldots} \sum_{k=0}^{\infty} P_0 \ldots P_{k-1} y_{n-1} \geq$$

$$\geq \sup_{P_0, \ldots} \sum_{k=0}^{\infty} P_0 \ldots P_{k-1} \sup_{x, l} \sum_{\ell=0}^{\infty} \binom{k+n-2}{k} P_k \ldots P_{k+\ell-1} |r_{P_k+\ell}|$$

$$= \sup_{P_0, \ldots} \sum_{m=0}^{\infty} \binom{k+n-2}{k} P_0 \ldots P_{m-1} |r_{P_m}|$$

$$= \sup_{P_0, \ldots} \sum_{m=0}^{\infty} \binom{m+n-1}{m} P_0 \ldots P_{m-1} |r_{P_m}|$$

So $y_n < \infty$ implies $z_a < \infty$ for $a_k(i) = O(k^{n-1})$, $k \to \infty$. The converse is not true, as shown by the following example.

**Counterexample 3.2.**

The states 1, 2, ... are absorbing with reward 0. In the states $n'$, $n = 1, 2, \ldots$, there are two actions. Action 1 yields reward 0 and a transition to state $(n+1)'$ action 2 yields reward $n^{-1}$ and a transition to state $n$. Obviously we have for all $R \in M$

$$\mathbb{E}_R \sum_{n=0}^{\infty} (n+1) |r(X_n, A_n)| \leq 1$$
but since $y_1(n') = n^{-1}$ we have for the strategy $R^*$ yielding transitions from $n'$ to $(n + 1)'$ etc. that

$$\mathbb{E}_{1,R^*} \sum_{n=0}^{\infty} y_1(X_n) = \infty.$$ 

But if we make a slightly stronger assumption then

$$\sup_{R \in \mathcal{M}} \sum_{n=0}^{\infty} n^{N-1} \mathbb{E}_{R^*} |r(X_n, A_n)| < \infty$$

the finiteness of the functions $y_1, \ldots, y_N$ defined in 1.7 can be shown.

**Theorem 3.3.**

If for a nondecreasing sequence of numbers $a_0, a_1, \ldots$, with $a_n \in \mathbb{R}$ and

$$b := \sum_{n=0}^{\infty} a_n^{-1} < \infty$$

it holds that

$$u := \sup_{R \in \mathcal{M}} \mathbb{E}_{R^*} \sum_{n=0}^{\infty} a_n^{N-1} |r(X_n, A_n)| < \infty$$

then the functions $y_1, \ldots, y_N$ defined in 1.7 are finite and satisfy the inequalities $y_k \leq ub^{k-1} a_0^{-k+N}$, $k = 1, \ldots, N$.

**Proof.** We will prove by induction

$$\sup_{R \in \mathcal{M}} \mathbb{E}_{R^*} y_k(X_n) \leq ub^{k-1} a_0^{-k+N} a_n^{-N}$$

for $k = 1, 2, \ldots, N - 1$, $n = 0, 1, 2, \ldots$.

Set $k = 1$. Using $y_1 = z$ (by definition) and

$$\sup_{R \in \mathcal{M}} \mathbb{E}_{R^*} z(X_n) \leq u a_n^{-1-N}$$

(from lemma 2.2 and theorem 2.1.ii) we get

$$\sup_{R \in \mathcal{M}} \mathbb{E}_{R^*} y_1(X_n) \leq u a_n^{-1-N}$$

for $n = 0, 1, 2, \ldots$.
Now let us assume
\[ \sup_{R \in M} E R^n k (X_n) \leq u_k^{-1} a^{-N} \] for \( k = 1, \ldots, m \leq N - 2 \) and \( n = 0, 1, \ldots \)
and prove that the inequalities hold for \( k = m + 1 \).

\[ \sup_{R \in M} E R^{m+1} (X_n) = \sup_{R \in M} P_0 \ldots P_{n-1} \sup_{R \in M} \sum_{k=0}^{\infty} \tilde{P}_0 \ldots \tilde{P}_{k-1} y_m \]

\[ = \sup_{P_0 \ldots P_{n-1}} \sup_{P_0 \ldots P_{n-1}} \sum_{k=0}^{\infty} P_0 \ldots P_{n-1} \tilde{P}_0 \ldots \tilde{P}_{k-1} y_m \]

\[ \leq u_k^{m-1} \sum_{n+\ell}^{m+1-N} a^{-N} a^{-\ell} \leq u_k^{-1} a^{-N} \] 

Thus we proved
\[ \sup_{R \in M} E R^n k (X_n) \leq u_k^{-1} a^{-N} \] for \( k = 1, 2, \ldots, N - 1, n = 0, 1, \ldots \).

Setting \( n = 0 \) we get \( y_k \leq u_k^{-1} a^{-N} \), \( k = 1, \ldots, N - 1 \) and with

\[ y_N = \sup_{R \in M} E R \sum_{n=0}^{\infty} y_{n+1} (X_n) \]

we get \( y_N \leq u_k^{-1} N \). (And obviously \( y_1, \ldots, y_N \) are finite).

**Corollary 3.4.**

If \( a_n = n^{-k+\varepsilon} \) for \( n = 0, 1, \ldots \), and some \( \varepsilon > 0 \), then \( z_a < \infty \) implies the existence of (finite) Liapunov functions \( \varepsilon_1, \ldots, \varepsilon_{k+1} \) satisfying 3.1.

This is immediate from theorem 3.3 with

\[ \sum_{n=0}^{\infty} n^{1+\varepsilon/k} < \infty . \]
4. Liapunov Functions and Successive Approximations

In this section we first formulate sufficient conditions for stability in terms of Liapunov functions $\ell_1$ and $\ell_2$ (of order 1 and order 2 respectively).

Lemma 4.1.

If some Liapunov function $\ell_1$ (of order 1) exists and if in addition

$$\lim_{n \to \infty} U_n \ell_1 = 0$$

then the problem is stable with respect to scrap functions $s \in V(\ell_1)$.

**Proof.** Since $z \leq \ell_1$ we have

$$\lim_{n \to \infty} U_n z = 0.$$ 

By lemma 2.2, theorems 2.4 and 2.3 we have the desired result.

Lemma 4.2.

If Liapunov functions $\ell_1$ and $\ell_2$ exist, then

$$\lim_{n \to \infty} U_n \ell_1 = 0.$$ 

**Proof.** Consider a new reward structure: $\tilde{r}_p := r_p - P \ell_1, \ P \in P$.

For all $R \in M$ we have

$$\ell_1 = \sum_{n=0}^{\infty} P_0 \cdots P_{n-1} \tilde{r}_p + \lim_{n \to \infty} P_0 \cdots P_n \ell_1.$$ 

Since

$$\ell_2 \geq \sum_{n=0}^{\infty} P_0 \cdots P_{n-1} \ell_1 \text{ for all } R \in M,$$

we have

$$\lim_{n \to \infty} P_0 \cdots P_n \ell_1 = 0.$$ 

Hence $\ell_1$ is the function $y_1$, defined in 1.7, for this new model. Therefore,
by theorem 3.1, lemma 2.2 and theorem 2.1 we have the desired result.

As a direct consequence of lemma's 4.1 and 4.2 we have

**Theorem 4.3.**

If Liapunov functions $\ell_1$ and $\ell_2$ exist, then the problem is stable with respect to scrapfunctions $s \in V(\ell_1)$.

We note that sometimes Liapunov functions $\ell_1$ and $\ell_2$ can be found rather simple, while $y_1$ and $y_2$ are difficult to obtain.

**Remark 4.4.**

If we assume besides the existence of a first order Liapunov function $\ell_1$, the compactness of $P$ and the continuity of $P\ell_1$, as function of $P$, then a sufficient condition for

$$\lim_{n \to \infty} U^n \ell_1 = 0$$

is

$$\lim_{n \to \infty} P^n \ell_1 = 0 \quad \text{for all } P \in P.$$  

The proof of this statement proceeds in a way similar to the proof of lemma 5.7 in [13].

**Theorem 4.5.**

Let $\ell_1$ and $\ell_2$ be Liapunov functions (of order 1 and 2 respectively) and define for a function $s \in V(\ell_1)$

$$b_1 := \inf \{ \ell_1^{-1}(i)(Us - s)(i) \mid i \in E, \ell_1(i) > 0 \}$$

$$b_2 := \sup \{ \ell_1^{-1}(i)(Us - s)(i) \mid i \in E, \ell_1(i) > 0 \}$$

then

$$(4.1) \quad s - b_1 \ell_2 \leq v \leq s + b_2 \ell_2.$$
Proof. First observe that \( s \in V(\ell_1) \) then also \( U_s \in V(\ell_1) \) so the set \( \{ i \mid \ell_1(i) = 0 \} \) gives no trouble. Since \( U_s \leq s + b_2 \ell_1 \) and \( \ell_1 \leq \ell_2 \) we have

\[
U^2_s \leq \sup_p \{ r_p + P(s + b_2 \ell_1) \} \leq U_s + b_2 \ell_1 \leq U_s + b_2 \ell_2
\]

Similarly, from \( U^k_s \leq s + b_2^+ \ell_2 \) it follows that \( U^{k+1}_s \leq s + b_2^+ \ell_2 \), hence \( U^n_s \leq s + b_2^+ \ell_2 \) for \( n = 1, 2, \ldots \). Since the problem is stable (theorem 4.3) we have

\[
v = \lim_{n \to \infty} U^n_s \leq s + b_2^+ \ell_2.
\]

The proof of the left inequality is similar. \( \square \)

The following, somewhat weaker, but more elegant inequality is now immediate.

\[
(4.2) \quad \| v - s \|_{\ell_2} \leq \| U_s - s \|_{\ell_1}.
\]

Remark.

If we have functions \( \ell_1 \) and \( \ell_2 \) satisfying the inequalities 3.1 but \( \ell_2(i) = \infty \) for some \( i \) then we may separate the state space into \( E_1 := \{ i \in E \mid \ell_2(i) < \infty \} \) and \( E_2 := E \setminus E_1 \). Since \( \ell_2(i) < \infty \) implies \( \ell_2(j) < \infty \) for all \( j \in E \) which can be reached under some strategy from state \( i \), we have that \( \ell_1 \) and \( \ell_2 \) are Liapunov functions on the smaller model with state space \( E_1 \). Hence all results can be generalized to that situation.

If for some \( P, r_p + P s = U_s, \| U_s - s \|_{\ell_1} \) is small and \( \ell_2 < \infty \) one may use the stationary strategy \( R := (P, P, \ldots) \). In section 7, th. 7.2 we give bounds for the value of this strategy.

It is well-known that the \( \beta \)-discounted dynamic programming model

\[
( \sum_{j \in E} p(j \mid i, a) \leq \beta < 1 \text{ for all } i \in E \text{ and } |r_p| \leq M \text{ for some } M \in \mathbb{R} \text{ and all } P \in \mathcal{P} )
\]
can be brought into our framework by defining an extra absorbing state \(-1\) with \(r(-1,a) = 0\) for all \(a \in A\) and
\[
p(-1 \mid i,a) = 1 - \sum_{j \in E} p(j \mid i,a), \quad i \in E.
\]

In this new model we can take as Lyapunov functions the functions defined by \(\ell_k(i) = M(1 - \beta)^{-k}, i \in E\), \(\ell_k(-1) = 0\) \(k = 1,2\) and then 4.1 becomes slightly weaker than the MacQueen bounds [19] since we work with \(b^-_1\) and \(b^+_2\) instead of \(b_1^-\) and \(b_2^+\).

In the following theorem \(s\) is an approximation for \(v\) with known bounds \(b^-_1\) and \(b^+_2\). At the price of extra calculation of \(\overline{u}(b^-_1)\) and \(\overline{u}(b^+_2)\) we obtain bounds for \(v^s_n\).

**Theorem 4.6.**

If \(s - b^-_1 \leq v \leq s + b^+_2\) then \(v^s_n - \overline{u}^n b^-_1 \leq v \leq v^s_n + \overline{u}^n b^+_2\).

**Proof.** For \(n = 1\) the statement is trivial. Suppose it holds for \(n = k\). Then
\[
v - v^s_{k+1} \leq \sup_{p} \{r_p + P v\} - \sup_{p} \{r_p + P v^s_k\} \leq \sup_{p} \overline{u}^k b^+_2 = \overline{u}^{k+1} b^+_2
\]
and
\[
v^s_{k+1} - v \leq \sup_{p} \{r_p + P v^s_k\} - \sup_{p} \{r_p + P v\} \leq \overline{u}^{k+1} b^-_1
\]
If there is a sequence \(P_1, P_2, \ldots\) such that
\[
\sup_{p} \{r_p + P v^s_n\} = r_p + P_{n+1} v^s_n \quad \text{for } n = 1, 2, \ldots
\]
then we can use \(P_{n-1} \ldots P_1 b^-_1\) instead of \(\overline{u}^n b^-_1\).

Note that we may choose \(b^+_2 = 0\) if \(s \geq v\).

Finally we can use these bounds to eliminate suboptimal actions. (We use the notation with explicitly written actions \(a\)).

Action \(a\) is called **suboptimal** or nonconserving if
\[
r(i,a) + \sum_{j} p(j \mid i,a)v(j) < \sup_{a \in A} \{r(i,a) + \sum_{j} P(j \mid i,a)v(j)\}
\]
Hence if $b_1$ and $b_2$ are bounds on $v$, $b_1 \leq v \leq b_2$ it holds that action $a$ is suboptimal if

$$r(i,a) + \sum_{j \in E} p(j \mid i,a) b_2 < \sup_{a \in A} \left\{ r(i,a) + \sum_{j \in E} P(j \mid i,a) b_1 \right\}.$$ 

In theorem 4.7 we prove that elimination of suboptimal actions gives a new model with the same value function. We only assume the model satisfies some strong convergence condition.

In [14] a similar property is proved without this condition.

**Theorem 4.7.**

Suppose that some strong convergence condition holds. Consider a new model with $\tilde{P} \subseteq P$ such that for all $\epsilon > 0$ there is a $P \in \tilde{P}$ with $r_P + P v \geq v - \epsilon$. Then the new model has the same value function.

**Proof.** Fix $\epsilon > 0$, let $\epsilon_n := \epsilon.2^{-(n+1)}$ and choose $P_n \in \tilde{P}$ such that

$$r_{P_n} + \epsilon_n + P_{n} v \geq v.$$ 

Iteration of this inequality yields

$$\sum_{n=0}^{N} P_0 \cdots P_{n-1}(r_{P_n} + \epsilon_n) + P_0 \cdots P_N v \geq v.$$ 

Hence

$$\sum_{n=0}^{\infty} P_0 \cdots P_{n-1} r_{P_n} + \sum_{n=0}^{\infty} \epsilon_n \geq v$$

since by the strong convergence condition

$$\lim_{n \to \infty} P_0 \cdots P_n v = 0.$$ 

Therefore

$$\sum_{n=0}^{\infty} P_0 \cdots P_{n-1} r_{P_n} \geq v - \epsilon,$$

consequently the supremum in this model equals $v$. $\square$
As in [7] and [8] we can also exclude actions for a finite number of iterations instead of all future iterations.

Fix some scrapfunction $s$. For notational convenience we omit the dependence on $s$ in the following definitions:

$$v_n(i,a) := r(i,a) + \sum_{j \in E} p(j | i,a) v_{n-1}^s(j)$$

$$d_n(i,a) := v_n^s(i) - v_n(i,a)$$

$$b_{1,n} := \inf_{i \in E} \{v_n^s(i) - v_{n-1}^s(i)\}, \quad b_{2,n} := \sup_{i \in E} \{v_n^s(i) - v_{n-1}^s(i)\}$$

$$\phi_n := b_{2,n} - b_{1,n}$$

**Theorem 4.8.**

(i) \(d_{n+k+1}(i,a) \geq d_n(i,a) - \sum_{\ell=0}^{k} \phi_{n+k+\ell}\)

(ii) if \(d_n(i,a) - \sum_{\ell=0}^{k} \phi_{n+k+\ell} > 0\) then action $a$ is suboptimal at stage $n+k+1$.

**Proof.** (ii) is a direct consequence of (i). Since

$$v_{n+1}(i,a) - v_n(i,a) = \sum_{j \in E} p(j | i,a) \{v_n^s(j) - v_{n-1}^s(j)\} \leq b_{2,n}$$

and

$$v_{n+1}^s(i) - v_n^s(i) \geq \inf_{a \in A} \sum_{j \in E} p(j | i,a) \{v_n^s(j) - v_{n-1}^s(j)\} \geq b_{1,n}$$

we have by subtraction of these inequalities:

$$d_{n+1}(i,a) = v_{n+1}^s(i) - v_{n+1}(i,a) \geq d_n(i,a) - \phi_n.$$ 

Iteration of this inequality yields the desired result. \(\square\)

Hence, if we determine at stage $n$: $d_n(i,a)$ and at each following stage: $\phi_{n+k}$, we need not compute $v_{n+k+1}(i,a)$ as long as

$$d_n(i,a) - \sum_{\ell=0}^{k} \phi_{n+k+\ell} > 0.$$
In this section we show how the contracting dynamic programming model introduced by Van Nunen [20] fits into the framework of strong convergence and Liapunov functions. The model assumptions are as follows:

There exist a finite function \( b \) and a bounding function \( \mu \) and there are constants \( k, k' > 0 \) and \( \rho, \rho' \) with \( 0 \leq \rho, \rho' < 1 \), such that

\[
(i) \quad \sup_{\mathbb{R}} \sum_{n=0}^{\infty} \mathbb{E}_{\mathbb{R}}[|b(X_n)|] < \infty
\]

and for all \( P \in \mathcal{P} \)

\[
(ii) \quad \| r - b \|_{\mu} \leq k
\]

\[
(iii) \quad P\mu \leq \rho \mu
\]

\[
(iv) \quad \| P - \rho b \|_{\mu} \leq k'.
\]

In the papers of Shapley [22], Blackwell [1] and Denardo [3] it is assumed that the rewards are bounded and that the operator \( U \) (def. 1.10) is a contraction with respect to the supremum norm. Veinott [23] showed that transient models can be transformed into discounted models using a similarity transformation which is equivalent to working with a bounding function (see below). Harrison [6] noticed that in many practical models with a countable state space the reward function is unbounded and he suggested a modification: he introduced the translation function \( b \). But he worked with \( \mu \equiv 1 \). Lippman [17, 18] remarked that Harrison's model is too restrictive to include for example the M/M/1 queueing system with quadratic cost. He introduced a special bounding function: a polynomial. Wijngaard [25] considered exponential bounding functions to study inventory models with the average cost criterion. Wessels [24] gave the first systematic treatment of general bounding functions for total return models with a countable state space. Van Hee and Wessels [11] studied necessary and sufficient conditions for the existence of a bounding function \( \mu \) such that for all \( P \in \mathcal{P} \): \( P\mu \leq \rho \mu \), \( 0 \leq \rho < 1 \). Hinderer [12] used bounding functions for finite stage dynamic programming models with a general state space.

We shall consider the contracting dynamic programming model in more detail.
Let us denote
\[ w_p := (1 - \rho)^{-1} (b - Pb) \]
then by iteration, we find:
\[
\sum_{n=0}^{N} P_0 \ldots P_{n-1} w_p P_n + P_0 \ldots P_N (1 - \rho)^{-1} b = (1 - \rho)^{-1} b
\]
Since by 5.1 i)
\[
\lim_{N \to \infty} P_0 \ldots P_N b = 0
\]
we have
\[
\sum_{n=0}^{\infty} P_0 \ldots P_{n-1} w_p P_n = (1 - \rho)^{-1} b .
\]
Hence the dynamic programming model with reward function
\[
\tilde{r}_p := r_p - w_p , \quad P \in P
\]
is equivalent to the original problem.
However
\[
\|\tilde{r}_P\|_\mu < \infty .
\]
Indeed with 5.1 ii) and iv) we find
\[
\|\tilde{r}_p\|_\mu = \|r_p - w_p\|_\mu = \|r_p - (1 - \rho)^{-1} b + (1 - \rho)^{-1} Pb\|_\mu = (1 - \rho)^{-1} \| (1 - \rho) r_p - b + Pb\|_\mu = (1 - \rho)^{-1} \| (1 - \rho) (r_p - b) - \rho b + Pb\|_\mu \leq (1 - \rho)^{-1} \{(1 - \rho) \| r_p - b\|_\mu + \| Pb - \rho b\|_\mu \} < \infty .
\]
Hence the contracting dynamic programming model is equivalent to a model satisfying for \( P \in P \) and some \( k > 0 \):
(i) \( P \mu \leq \rho \mu \)
(ii) \( \| r_P \|_\mu \leq k \)
Note that this model can be reduced in a similar way to a discounted dynamic programming model by the transformations:
This is in fact the similarity transformation studied by Veinott [23].

From 5.2 i) and ii) we have immediately

\[ \sup_{\mathbb{R}} \mathbb{E}[|r(X_n, Y_n)|] \leq \sup_{\mathbb{R}} \prod_{i=0}^{n-1} |r_{i+1}| \leq k_0 \mu \]

and therefore, we have for \( 1 < \lambda < \rho^{-1} \)

\[ \sup_{\mathbb{R}} \mathbb{E}[\sum_{n=0}^{\infty} \lambda^n |r(X_n, Y_n)|] \leq k(1 - \lambda \rho)^{-1} \mu < \infty . \]

Thus the contracting dynamic programming model satisfies the strong convergence condition for the sequence \( a_n \equiv \lambda^n \). And since \( n^k = O(\lambda^n) (n \to \infty) \) for all \( k = 1, 2, \ldots \) we have by corollary 3.4 that there exist Liapunov functions \( \mathcal{L}_k \) satisfying 3.1 for \( k = 1, 2, \ldots \).

Apart from this one immediately sees that

\[ \mu + (1 - \rho)^{-1} P \mu \leq (1 - \rho)^{-1} \mu \]

thus

\[ |r_p| + k(1 - \rho)^{-1} P \mu \leq k(1 - \rho)^{-1} \mu . \]

Hence \( k(1 - \rho)^{-1} \mu \) suffices as Liapunov function \( \mathcal{L}_1 \), and it is easily checked that \( k(1 - \rho)^{-n} \mu, n = 1, 2, \ldots \) is a system of Liapunov functions satisfying 3.1.
6. Waiting Line Model with Controllable Input; an Example which is Strongly Convergent but not necessarily Contracting

In this section we consider as an example the waiting line model with controllable input which was studied in Chapter 5 in [13] and in [15]. In this queuing model the arrival process is Poisson with expected number of arrivals per unit time $\lambda_a$ where $a$ denotes the service cost. We assume that we can control the arrival process by choosing $a$ from the interval $[a_1, a_2]$. And we make the reasonable assumption that $\lambda_a$ decreases as $a$ increases. The service time distribution $F$ is general.

At each time a customer completes service, the service cost may be changed. We will be looking at the embedded Markov chain.

The states space becomes $E = \{0, 1, \ldots\}$ and the transition probabilities satisfy

$$ p(j \mid i, a) = \begin{cases} 0 & \text{if } j < i - 1 \\ k_{j-i+1}(a) & \text{if } j \geq i - 1 \end{cases} $$

with

$$ k_r(a) = \int_{0}^{\infty} e^{-\lambda_a s} r r(s)^{-1} \beta(s) \, dF(s). $$

Furthermore we assume

$$ \int_{1}^{\infty} s dF(s) < 1 $$

and $r(i, a) \geq \delta > 0$ for $i = 1, 2, \ldots$ and all $a \in A := [a_1, a_2]$

If one is looking for an average optimal strategy for this problem then one is interested in the behaviour of the system up to the first time the system empties again.

In order to study the behaviour until this time we modify the transition probabilities and rewards in state 0 as follows

$$ p(j \mid 0, a) := \delta_{0j} \quad \text{and} \quad r(0, a) := 0. $$
If this model is contracting then there exists a bounding function \( \mu \) satisfying

\[(i) \quad |r_p| \leq k \mu \text{ for some } k \in \mathbb{R} \text{ over all } P \in P \]

\[(ii) \quad P \mu \leq \rho \mu, \text{ for some } 0 \leq \rho < 1 \text{ and all } P \in P. \]

Now (i) implies \(|r(i,a)| \leq k \mu(i)\) and with \(r(i,a) \geq \delta > 0\) follows \(\mu(i) \geq \delta^{-1}, \; i \geq 1\). Now we may use theorem 2 in [11] which states that there exists a function \( \mu \) satisfying (ii) and

\[\inf_{i \geq 1} \mu(i) > 0\]

if and only if the lifetime \( N \) of the process (here the number of transitions until state 0 is reached) is exponentially bounded.

So in order that this model is contracting at least all moments of the life time must be finite and with the inequality

\[\mathbb{E}[N(N-1)\ldots(N-k+1)] = \sum_{\ell=k}^{\infty} \int_{0}^{\infty} \frac{\lambda^\ell}{\ell!} e^{-\lambda s} \ell(\ell-1)\ldots(\ell-k+1) dF(s) = \lambda^k \int_{0}^{\infty} s^k dF(s)\]

(cf. [15]) we see that all moments of the service time must be finite as well. Hence the model is certainly not contracting if not all moments of the service time are finite.

On the other hand it is shown in [15] that if the \( k \)-th moment of the service time is finite and if

\[\sup_{a} |r(i,a)| \leq A\ell^k\]

for some \( A \in \mathbb{R} \) and all \( i \in S \) then there exist Liapunov functions \( y_1, \ldots, y_{k-\ell}, \ell < k \). We will prove this here using a completely different approach.

First one may show that if the \( k \)-th moment of the service time is finite then also the \( k \)-th moment of the lifetime of the imbedded process is finite. This may be seen as follows.

It is clear that the lifetime is maximized if we use the strategy \( \hat{R} \) which corresponds to the minimal service cost in each state. For that strategy we have an \( M|G|1 \) queue. And the lifetime of the imbedded process is now equal to the number of customers \( N \) in the busy period of the \( M|G|1 \) queue.

Let \( F^* \) be the Laplace transform of the service time and \( N^* \) the transform of the distribution of the number of customers in a busy period. Then we have the following relation between \( F^* \) and \( N^* \)
\[ N^*(t) = e^{-tF^*(\lambda - \lambda N^*(t))}, \quad t > 0 \]

where \( \lambda \) is the Poisson parameter (cf. Cohen [2] p. 250). Differentiating this equation once with respect to \( t \) gives

\[ (6.1) \quad N^{*'}(t) = \frac{-e^{-tF^*(\lambda - \lambda N^*(t))}}{1 + \lambda F^{*'}(\lambda - \lambda N^*(t))} \]

The denominator is bounded from below by

\[ 1 - \int_0^\infty (\lambda - \lambda s) dF(s) > 0. \]

It is well-known (see for example Feller [5] p. 412) that \( N^{*k}(t) \) has a finite limit for \( t \to 0 \) iff

\[ \sum_{n=0}^{\infty} n^k P(N = n) < \infty. \]

Then

\[ \sum_{n=0}^{\infty} n^k P(N = n) = (-1)^k N^{*k}(0). \]

Differentiating (6.1) one may show by induction that if \( F^{*k}(t) \) has a finite limit for \( t \to 0 \) for \( k = 1, \ldots, k \) then \( N^{*k}(t) \) has a finite limit for \( t \to 0 \) as well.

So we conclude that if the \( k \)-th moment of the service time is finite then also the \( k \)-th moment of the lifetime of the embedded process is finite.

Now suppose

\[ \mathbb{E} R^{N^k} < \infty \quad \text{and} \quad \sup_P |r_P(i)| \leq A_i^{k-m-1} \quad \text{for some} \ A \in \mathbb{R} \quad \text{and all} \ i \in \mathbb{E}. \]

Then we have for all \( R \)

\[ \mathbb{E} R^{\sum_{x=0}^{\infty} x^m |r(X_x, Y_x)|} = \sum_{t=0}^{\infty} \mathbb{P}_R (N = t) \sum_{x=0}^{t} \mathbb{E}_R [r(X_x, Y_x) | N = t] \]

\leq \sum_{t=0}^{\infty} \mathbb{P}_R (N = t) \sum_{x=1}^{t} t^{m-1} A t^{m-k-1} \]

\[ = A \sum_{t=0}^{\infty} t^k \mathbb{P}_R (N = t) \leq A \sum_{t=0}^{\infty} t^k \mathbb{P}_R (N = t) < \infty. \]
Where the inequality
\[
\sum_{t=0}^{\infty} \mathbb{E}_R[|r(x_t, y_t) | N = t] \leq A t^{k-m-1}
\]
follows immediately from the fact that in the embedded process only one customer is served per unit of time.
So we see that
\[
\int_0^\infty s^k F(s) < \infty \text{ and } \sup_p | r_p(i) | \leq A i^{k-m-1}
\]
for some \( A \in \mathbb{R} \) and all \( i \in E \) imply, using corollary 3.4, the finiteness of the functions \( y_1, \ldots, y_m \).
Reasoning in a similar way one may show that for \( m = 0 \) the model is strongly convergent (and thus \( y_1 < \infty \)).
7. Nearly Optimal Strategies

In this section we collect some results with respect to nearly optimal strategies for the strongly convergent case. But before we do so we first give an example which shows that there need not exist for all \( \varepsilon > 0 \) a stationary strategy \( P^{(\infty)} \) satisfying

\[
(7.1.) \quad v(\cdot, P^{(\infty)}) \geq v - \varepsilon(1 + |v|) \varepsilon
\]

if we only assume

\[
\sup_{\mathcal{R}} \mathbb{E}_{R} \sum_{n=0}^{\infty} |r(X_n, Y_n)| < \infty
\]

but not 1.8, the uniform tail property or positivity of all \( r(i,a) \). For the positive case Ornstein [21] proved the existence of a \( P^{(\infty)} \) satisfying 7.1.

Example 7.1.

In state \( n \) there are two actions. Action 1 gives reward 0 and a transition to state \( n + 1 \) with probability

\[
a_n = \frac{\frac{1}{2}}{b_n},
\]

where

\[
b_n = 1 + \frac{1}{n},
\]

and with probability \( 1 - a_n \) the system leaves \( E \). Action 2 gives a reward \( 2^n \) and the system leaves \( E \) with probability 1.

\( v \) may be found as follows

\[
v(n) = \sup (2^n, a_n 2^{n+1}, a_n a_n 2^{n+2}, \ldots) = 2^n \sup(1, \frac{b_n}{b_{n+1}}, \frac{b_n}{b_{n+2}}, \ldots)
\]

\[
= 2^n b_n = 2^n(1 + \frac{1}{n}), \text{ since } b_n + 1 \text{ as } n \to \infty.
\]
Thus \( v(n') = 1 - 2^n(1 + \frac{1}{n}) + 2^n(1 + \frac{1}{n}) = 1 \).

We will show that there does not exist a stationary strategy \( P^{(\omega)} \) for which
\[ v(n', P^{(\omega)}) \geq 0 \] for all \( n = 1, 2, \ldots \).

Any stationary strategy may be characterized by the probabilities \( y_n \) by
which action 2 is taken in state \( n, \ n = 1, 2, \ldots \). (We consider randomized
strategies since when we were looking for an example we have seen that it
may occur that though there is no pure \( \varepsilon \)-optimal strategy there does exist
a randomized one).

We see that for this strategy
\[ v(n', R) \leq 1 - 2^n(1 + \frac{1}{n}) + y_n 2^n + (1 - y_n) 2^n(1 + \frac{1}{n}). \]

So strategy \( R \) gives for state \( n' \) an immediate loss of \( y_n 2^n/n \) compared to what
could be gained. In order that this loss is smaller than 1 we must have
\( y_n \leq n2^{-n} \). Now let us consider an arbitrary strategy \( R \) with \( y_n \leq n2^{-n} \) for
all \( n \) and see what its total expected reward for state \( n \) is. Using the
inequalities \( \alpha_n \leq 2/3, 1 - y_n \leq 1 \) and \( y_n \leq n2^{-n}, \ n = 1, 2, \ldots \) we get
\[ v(n, R) = 2^n y_n + \alpha_n (1 - y_n) y_{n+1} 2^{n+1} + \alpha_n y_n \alpha_n (1 - y_n) (1 - y_{n+1}) y_{n+2} 2^{n+2} + \ldots \]
\[ \leq 2^n n2^{-n} + \frac{2}{3}(n + 1)2^{-n+1} + \frac{2}{3}(n + 2)2^{-n+2} \ldots \]
\[ = n + \frac{2}{3}(n + 1) + \frac{4}{9}(n + 2) + \ldots = 3n + 6. \]

So for \( n \geq 4 \) \( v(n, R) \leq -1 \). Hence no stationary strategy \( P^{(\omega)} \) exists
with \( v(i, R) \geq v(i) - \varepsilon(1 + |v(i)|) \varepsilon \) for all \( \varepsilon < \frac{1}{2} \). This concludes our
counterexample.

Now we continue with some positive results.

If the model is strongly convergent then Howards' policy iteration algorithm
converges. And as a result we conclude that in the strong convergence case
it holds that for all \( i \in E \) and all \( \varepsilon > 0 \) there exists a stationary strategy
\( P^{(\omega)} \) such that
\[ v(i, P^{(\omega)}) \geq v(i) - \varepsilon. \] (cf. [10]).

The following results deal with uniform \( \varepsilon \)-optimality on \( E \), in some sense.
If for a sequence $a = (a_0, a_1, \ldots)$ with $a_n \to \infty$ uniformly on $E$ it holds that $z_a < \infty$, then for all $\varepsilon > 0$ there exists a stationary strategy $p^{(\infty)}$ such that

$$v(\cdot, p^{(\infty)}) \geq v - \varepsilon z_a$$

(cf. [10]).

And if $w/a \to 0$ uniformly on $E$ then there exists for all $\varepsilon > 0$ a stationary strategy $p^{(\infty)}$ satisfying $v(\cdot, p^{(\infty)}) \geq v - \varepsilon e$ (cf. [10]).

Theorem 7.2. Let $\ell_1$ and $\ell_2$ be Lyapunov functions of order 1 and 2 and let either $s \in V(\ell_1)$ or $\limsup T^{-1} s \leq 0$. If furthermore $r_p + P s \geq Us - \varepsilon \ell_1$, then

$$v(\cdot, p^{(\infty)}) \geq s - \varepsilon \ell_2.$$

Proof. Iterating $r_p + P s \geq s - \varepsilon \ell_1$ gives us

$$T^{-1} \sum_{n=0}^{T-1} P^n r_p + P^n s \geq s - \varepsilon \sum_{n=0}^{T-1} P^n \ell_1 \geq s - \varepsilon \ell_2.$$

Letting $T \to \infty$ yields the desired result.

Theorem 7.3.

If $z < \infty$ and

$$\sup_{P \in \mathcal{P}} \sum_{n=1}^{\infty} P P_{n-1} r_p < \infty$$

then there exists for any state $i \in S$ and for all $\varepsilon > 0$ a stationary strategy $p^{(\infty)}$ with

$$v(i, p^{(\infty)}) \geq v(i) - \varepsilon.$$

Proof. The proof proceeds analogous to the proof of theorem 13.6 in [13].
Fix \( i \in E \) and \( \varepsilon > 0 \). Let strategy \( R \) be such that \( v(i,R) \geq v(i) - \frac{\varepsilon}{4} \).

Choose \( 0 < \alpha < 1 \) such that

\[
E_{i,R} \sum_{n=0}^{\infty} \alpha^n r(X_n,Y_n) \geq v(i) - \frac{\varepsilon}{4}
\]

and

\[
(1 - \alpha) \sup_p \sum_{n=0}^{\infty} nP^n_{r_Q}(i) \leq \frac{\varepsilon}{2}.
\]

The \( \alpha \)-discounted problem is strongly convergent, hence by 7.2, there exists a \( Q \) such that

\[
\sum_{n=0}^{\infty} \alpha^n Q^n r_Q(i) \geq \sup_{\alpha} E_{i,R} \sum_{n=0}^{\infty} \alpha^n r(X_n,Y_n) - \frac{\varepsilon}{4} \geq v(i) - \frac{\varepsilon}{4}.
\]

Since \( 1 - \alpha^n \leq (1 - \alpha)n \) for \( 0 < \alpha < 1 \) and \( n = 0, 1, \ldots \) we have

\[
\sum_{n=0}^{\infty} Q^n r_Q(i) \geq \sum_{n=0}^{\infty} \alpha^n Q^n r_Q(i) - \sum_{n=0}^{\infty} (Q^n - \alpha^n Q^n) r_Q(i) \]

\[
\geq v(i) - \frac{\varepsilon}{2} - \sum_{n=0}^{\infty} (1 - \alpha) nQ^n r_Q(i) \geq v(i) - \varepsilon.
\]

Hence \( v(i,Q) \geq v(i) - \varepsilon \).

Finally a result on optimal strategies.

**Theorem 7.4.**

If the model is strongly convergent then any conserving \( P \), i.e. \( r_p + Pv = v \), constitutes a stationary optimal strategy.

**Proof.** Iterating \( r_p + Pv = v \) we get

\[
\sum_{n=0}^{N-1} p^n r_p + p^N v = v.
\]

Since

\[
\sum_{n=0}^{N-1} p^n r_p \rightarrow v(\cdot, p^{(\infty)}) \quad (N \rightarrow \infty)
\]

and \( p^N v \rightarrow 0 \quad (N \rightarrow \infty) \) (2.3) we have \( v(\cdot, p^{(\infty)}) = v \).
Hence if the model is strongly convergent, $P$ compact, $w < \infty$, $r$, and $Pw$ continuous of $P$ then there exists a stationary optimal strategy. Since with the compactness and continuity assumptions one may show the existence of a conserving $P$.

See also chapter 4 in [13].

References


[12] Hinderer, K., Bounds for stationary finite-stage dynamic programs with unbounded reward functions (To be published).


