Successive approximation for average reward Markov games

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Technology

PROBABILITY THEORY, STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum-COSOR 77-10

Successive approximations for average reward Markov games

by

J. van der Wal

Eindhoven, April 1977

The Netherlands

Successive approximations for average reward Markov games

by

J. van der Wal

Abstract. This paper considers two-person zero-sum Markov games with finitely many states and actions with the criterion of average reward per unit time. Two special situations are treated and it is shown that in both cases the method of successive approximations yields an ε-band for the value of the game as well as stationary ε-optimal strategies. In the first case all underlying Markov chains of pure stationary optimal strategies are assumed to be unichained. In the second case it is assumed that the functional equation $Uv = v + ge$ has a solution.

1. Introduction, Notations and some Preliminary Results

In this paper we deal with two aspects of average reward two-person zero-sum Markov games with finitely many states and actions. We will not go into the important and still unsolved question whether these games have a value within the class of all strategies. What we know is that in general they will neither have a value within the class of stationary strategies, nor in the class of Markov strategies (cf. Gillette [4], Blackwell and Ferguson [1]).
It has been shown by Gillette [4] and afterwards by Hoffman and Karp [5] that the game has a value within the class of stationary strategies if for each pair of stationary strategies the underlying Markov chain is irreducible. This condition has been weakened by Rogers [6] and Sobel [9] who still demanded that the underlying Markov chain was unichained but allowed for some transient states. Federgruen [3] has shown that the unichain restriction may be replaced by the condition that the underlying Markov chains corresponding to a pair of (pure) stationary strategies all have the same number of irreducible subchains.
Here we give two results obtained from a successive approximations approach. In section 2 we show that in the situation of Rogers and Sobel (one recurrent chain and possibly some transient states) successive approximations yield ε-optimal stationary strategies and an ε-band for the value of the game. In order to find them we first apply Schweitzer's [7] data transformation to obtain an equivalent Markov game. In section 3 we consider games that do have a

value independent of the starting state. Using again Schweitzer's transformation we show that successive approximations yield ε-optimal stationary strategies.

First we specify the game and give a number of notations. We are concerned with a dynamic system with finite state space $S := \{1,2,\ldots,N\}$ which is observed at times $t = 0,1,\ldots$. The behaviour of the system is influenced by two players, P_1 and P_2, having completely opposite aims. For each $i \in S$ two finite nonempty sets of actions exist, one for each player, denoted by K_i for P_1 and L_i for P_2. As a joint result of the state i and the two selected actions, k for P_1 and ℓ for P_2, the system moves to a new state j with probability $p(j|i,k,\ell)$, where $\sum_{j\in S} p(j|i,k,\ell) = 1$, and P_1 receives a (possibly negative) amount from P_2 denoted by $r(i,k,\ell)$.

Following Zachrisson [11] we shall call these two-person zero-sum games Markov games. Many authors, following Shapley [8], use the term stochastic games.

A strategy π for P_1 in this game is any function that specifies for each time $n = 0,1,\ldots$ and for each state $i \in S$ the probability $\pi(k|i,n,h_n)$ that action $k \in K_i$ will be taken as a function of i, n and the history h_n. By the history h_n up to time n we mean the sequence $h_n = (i_0,k_0,\ell_0,\ldots,i_{n-1},k_{n-1},\ell_{n-1})$ of prior states and actions (h_0 is the empty sequence). If all $\pi(k|i,n,h_n)$ are independent of h_n the strategy is called a Markov strategy, if moreover $\pi(k|i,n,h_n)$ does not depend on n the strategy is called stationary. A policy f for P_1 is any function such that f(i) is a probability distribution on K_i for all $i \in S$. Thus a Markov strategy prescribes a policy f_n for each time n and will be denoted by $(f_0,f_1,\ldots)$. And a stationary strategy is a Markov strategy with $f_n = f_0$, $n \geq 1$ and will be denoted by $f_0^{(\infty)}$.

Similarly we define strategies ρ and policies h for P_2.

We define $V_n(\pi,\rho)$ as the total expected reward vector for the first n periods when strategies π and ρ are used, and $g(\pi,\rho)$ by

$$g(\pi,\rho) = \liminf_{n\to\infty} n^{-1} V_n(\pi,\rho) \ .$$

We say that the game has a value G if

$$\inf_{\rho} \sup_{\pi} g(\pi,\rho) = \sup_{\pi} \inf_{\rho} g(\pi,\rho) = G \ .$$

(Instead of taking liminf in the definition we might just as well have taken limsup).

A strategy π_ε for P_1 is called ε-optimal, $\varepsilon \geq 0$, if

$$\inf_{\rho} g(\pi_\varepsilon,\rho) \geq G - \varepsilon e \quad (e^T = (1,\ldots,1))$$

and a strategy ρ_ε is called ε-optimal if

$$\sup_\pi g(\pi,\rho_\varepsilon) \le G + \varepsilon e \ .$$

A 0-optimal strategy is called optimal.

For convenience we define the vector $r(f,h)$ and matrix $P(f,h)$ by

$$r(f,h)(i) := \sum_{k\in K_i} \sum_{\ell\in L_i} f(i,k)h(i,\ell)[r(i,k,\ell)] \ ,$$

(where $f(i,k)[h(i,\ell)]$ denotes the probability by which action $k[\ell]$ is selected according to policy $f[h]$).

$$P(f,h)_{ij} := \sum_{k\in K_i} \sum_{\ell\in L_i} f(i,k)h(i,\ell)[p(j|i,k,\ell)] \ .$$

Since we will be engaged with successive approximations it is of notational convenience to introduce the operators $L(f,h)$ and U on $\mathbb{R}^N$ by

$$L(f,h)v := r(f,h) + P(f,h)v \ ,$$

$$Uv := \max_f \min_h L(f,h)v \ .$$

(maxmin is taken componentwise).
For a vector $w \in \mathbb{R}^N$ we define

$$\Delta w := \max_{i\in S} w(i), \ \nabla w := \min_{i\in S} w(i) \text{ and } sp(w) := \Delta w - \nabla w \ .$$

Now we start with some preliminary results.

Lemma 1. If a policy f_v satisfies for all h $L(f_v,h)v \ge Uv$ then we have

$$\inf_\rho g(f_v^{(\infty)},\rho) \ge \nabla(Uv - v).e$$

and similarly if h_v satisfies for all f $L(f,h_v)v \le Uv$ then

$$\sup_\pi g(\pi,h_v^{(\infty)}) \le \Delta(Uv - v)(i).e \ .$$

Proof. We only prove the first statement, the proof of the second is similar. Let P_1 play strategy $f_v^{(\infty)}$ then P_2 may restrict himself to stationary strategies. (This may be shown using a games version of Derman and Strauch's result [2], see for example the proof of theorem 2.3 in Stern [10]).
So we only have to show

$$\min_h \lim_{n\to\infty} n^{-1} V_n(f_v^{(\infty)}, h^{(\infty)}) \geq \nabla(Uv - v).e$$

or

$$\lim_{n\to\infty} n^{-1} L^n(f_v, h)0 \geq \nabla(Uv - v).e \quad \text{for all } h .$$

This follows immediately with

$$L^n(f_v,h)0 \geq L^n(f_v,h)v - \Delta v.e \geq L^{n-1}(f_v,h)v + \nabla(Uv - v).e - \Delta v.e$$

$$\geq \ldots \geq n.\nabla(Uv - v).e - \Delta v.e . \qquad \square$$

An immediate consequence of this lemma is the following

Corollary 1.

i) If $Uv - v = g.e$ for some $g \in \mathbb{R}$, then the game has value $g.e$ and the policies f_v and h_v from lemma 1 constitute stationary optimal strategies. I.e. if there exists a solution (v^*, g^*), $v^* \in \mathbb{R}^N$, $g^* \in \mathbb{R}$ of the functional equation

(1.1) $\qquad Uv - v + g.e$

then the game has value $g^*.e$ and strategies $f^{*(\infty)}$, $h^{*(\infty)}$ satisfying for all f and h

$$L(f^*,h)v^* \geq L(f^*,h^*)v^* (= Uv^*) \geq L(f,h^*)v^*$$

are optimal.

ii) If there exists a sequence $\{v_n\}$ of vectors in $\mathbb{R}^N$ with $sp(Uv_n - v_n) \to 0$ $(n \to \infty)$ then the game has a value and the strategy $f_n^{(\infty)}$ with $L(f_n,h)v_n \geq Uv_n$ for all h will be $sp(Uv_n - v_n)$-optimal.

Proof. i) is straightforward. The only difficulty in ii) is that we must show that there exists a $g^* \in \mathbb{R}$ such that $Uv_n - v_n \to g^*.e$ $(n \to \infty)$. But this is immediate once we realize that for any two vectors $v,w \in \mathbb{R}^N$

$$\nabla(Uv - v)(\leq g(f_v^{(\infty)}, h_w^{(\infty)})) \leq \Delta(Uw - w) . \qquad \square$$

Part ii) of corollary 1 is important if we deal with the method of successive approximations.

If we are going to apply the method of successive approximations then as in the case of Markov decision problems periodic behaviour will be unpleasant. In order to eliminate any periodicity one may apply Scheitzer's transformation [7].
Before we proceed to section 2 we will consider this transformation and we will show that it gives rise to an equivalent problem.

Assume we have a specific Markov game with data $r(.,.,.)$ and $p(.|.,.,.)$.

Now we may transform the data using Schweitzer's transformation as follows $(0 < \alpha < 1)$

$$\hat{r}(.,.,.) := (1 - \alpha)r(.,.,.)$$

$$\hat{p}(j|i,.,.) := \alpha\delta_{ij} + (1 - \alpha)p(j|i,.,.) \ .$$

That these two games are (in many respects) equivalent follows from the following lemma. The operators $L(f,h)$ and U and the function $g(.,.)$ for the transformed problem are denoted by $\hat{L}$, $\hat{U}$ and $\hat{g}$.

Lemma 2.

i) $\hat{L}(f,h)v - v = (1 - \alpha)[L(f,h)v - v]$ for all f and h .

ii) $\hat{U}v - v = (1 - \alpha)(Uv - v)$.

Proof.

i) From $\hat{r}(f,h) = (1 - \alpha)r(f,h)$ and $\hat{P}(f,h) = \alpha I + (1 - \alpha)P(f,h)$ we have

$$\begin{aligned}\hat{L}(f,h)v - v &= \hat{r}(f,h) + \hat{P}(f,h)v - v \\ &= (1 - \alpha)r(f,h) + \alpha Iv + (1 - \alpha)P(f,h)v - v \\ &= (1 - \alpha)[r(f,h) + P(f,h)v - v] \\ &= (1 - \alpha)[L(f,h)v - v] \ .\end{aligned}$$

ii) Since $1 - \alpha \geq 0$ the second statement follows now immediately. □

So if the functional equation of the original problem has a solution, say g^*, v^*, then the functional equation of the transformed problem $\hat{U}v = v + g.e$ has the solution $(1 - \alpha)g^*, v^*$. Similarly, if $\hat{U}\hat{v} = \hat{v} + \hat{g}.e$ then also $U\hat{v} = \hat{v} + (1 - \alpha)^{-1}\hat{g}.e$.

And if for example for all h $\hat{L}(\hat{f},h)v - v \geq \hat{g}e - \varepsilon e$, which implies by lemma 1 $\hat{g}(\hat{f}^{(\infty)},h^{(\infty)}) \geq \hat{g}e - \varepsilon e$, or $\hat{f}^{(\infty)}$ is ε-optimal in the transformed problem then $\hat{f}^{(\infty)}$ is also $(1-\alpha)^{-1}\varepsilon$-optimal in the original problem as follows from:

$$L(\hat{f},h)v - v = (1-\alpha)^{-1}(\hat{L}(\hat{f},h)v - v) \geq (1-\alpha)^{-1}\hat{g}e - (1-\alpha)^{-1}\varepsilon e =$$

$$= g^{*}e - (1-\alpha)^{-1}\varepsilon e \ .$$

Thus we see that both problems are equivalent with respect to those features which are important when we apply successive approximations. And therefore we will consider in the remainder of this paper only games satisfying for some $0 < \alpha < 1$ and all f and h

(1.2) $\qquad P(f,g) - \alpha I \geq 0 \ .$

2. The Unichained Model

In this section we consider Markov games satisfying the following assumption: Unichain assumption. For all pure stationary strategies the underlying Markov chain will consist of one recurrent subchain and possibly some transient states.

We will approximate the value (which is independent of the starting time) and find ε-optimal strategies by means of the method of successive approximations. We do this by showing that under the extra assumption $P(f,h) - \alpha I \geq 0$ for some $\alpha > 0$ and all f and h for any $v \in \mathbb{R}^N$ $sp(U^{n+1}v - U^{n}v)$ tends to zero geometrically as $n \to \infty$.

First we will derive some inequalities from which it will be clear that the main theorem of this section guarantees the convergence of successive approximations.

Let $v \in \mathbb{R}^N$ be given arbitrarily and let the policies f_k, h_k satisfy

$$L(f,h_k)U^{k-1}v \leq L(f_k,h_k)U^{k-1} \leq L(f_k,h)U^{k-1}v$$

for all policies f and h, $k = 1,2,\ldots$.

Then for all n

(2.1) $\qquad U^{n+1}v - U^{n}v = L(f_{n+1},h_{n+1})\ldots L(f_2,h_2)Uv - L(f_n,h_n)\ldots L(f_1,h_1)v$

$$\leq L(f_{n+1},h_n)\ldots L(f_2,h_1)Uv - L(f_{n+1},h_n)\ldots L(f_2,h_1)v$$

$$= P(f_{n+1},h_n)\ldots P(f_2,h_1)(Uv - v)$$

and similarly

$$(2.2) \qquad U^{n+1}v - U^{n}v \geq P(f_n,h_{n+1})\ldots P(f_1,h_2)(Uv - v) \ .$$

Let us denote for an arbitrary pair of strategies π,ρ the probability that at time n the system occupies state k given that the state at time 0 is i and strategies π and ρ are used by

$$\mathbb{P}_i^{\pi,\rho}(S_n = k) \ .$$

Let further $\pi_1[\rho_1]$ denote the strategy $(f_n,\ldots,f_2,f_1,f_1,\ldots)$ $[(h_n,\ldots,h_2,h_1,h_1,\ldots)]$ and $\pi_2[\rho_2]$ the strategy $(f_{n+1},\ldots,f_3,f_2,f_2,\ldots)$ $[(h_{n+1},\ldots,h_3,h_2,h_2,\ldots)]$.
Then for example

$$(P(f_n,h_{n+1})\ldots P(f_1,h_2))_{ij} = \mathbb{P}_i^{\pi_1,\rho_2}(S_n = j) \ .$$

Now we have for all $i,j \in S$

$$(2.3) \quad (U^{n+1}v - U^{n}v)(i) - (U^{n+1}v - U^{n}v)(j) \leq \sum_{k\in S} [\mathbb{P}_i^{\pi_2,\rho_1}(S_n=k) - \mathbb{P}_j^{\pi_1,\rho_2}(S_n=k)](Uv - v)(k)$$

$$= \sum_{k\in S} [\mathbb{P}_i^{\pi_2,\rho_1}(S_n=k) - \min\{\mathbb{P}_i^{\pi_2,\rho_1}(S_n=k),\mathbb{P}_j^{\pi_1,\rho_2}(S_n=k)\}](Uv - v)(k) +$$

$$- \sum_{k\in S} [\mathbb{P}_j^{\pi_1,\rho_2}(S_n=k) - \min\{\mathbb{P}_i^{\pi_2,\rho_1}(S_n=k),\mathbb{P}_j^{\pi_1,\rho_2}(S_n=k)\}](Uv - v)(k)$$

$$\leq \sum_{k\in S} [\mathbb{P}_i^{\pi_2,\rho_1}(S_n=k) - \min\{\mathbb{P}_i^{\pi_2,\rho_1}(S_n=k),\mathbb{P}_j^{\pi_1,\rho_2}(S_n=k)\}]\Delta(Uv - v) +$$

$$- \sum_{k\in S} [\mathbb{P}_j^{\pi_1,\rho_2}(S_n=k) - \min\{\mathbb{P}_i^{\pi_2,\rho_1}(S_n=k),\mathbb{P}_j^{\pi_1,\rho_2}(S_n=k)\}]\nabla(Uv - v)$$

$$= 1 - \sum_{k\in S} \min\{\mathbb{P}_i^{\pi_2,\rho_1}(S_n=k),\mathbb{P}_j^{\pi_1,\rho_2}(S_n=k)\}\,sp(Uv - v) \ .$$

Hence

$$sp(U^{n+1}v - U^{n}v) \leq 1 - \min_{i,j} \sum_{k\in S} \min\{\mathbb{P}_i^{\pi_2,\rho_1}(S_n=k),\mathbb{P}_j^{\pi_1,\rho_2}(S_n=k)\}\,sp(Uv - v) \ .$$

In theorem 2 we will prove that under the assumptions of this section there exists a $\gamma > 0$ such that for all π,ρ,π',ρ':

$$(2.4) \qquad \min_{i,j} \sum_{k\in S} \min\{\mathbb{P}_i^{\pi,\rho}(S_{N-1} = k), \mathbb{P}_j^{\pi',\rho'}(S_{N-1} = k)\} \geq \gamma .$$

Then we have from (2.3) and (2.4)

$$sp(U^N v - U^{N-1} v) \leq (1 - \gamma) sp(Uv - v) .$$

Thus $sp(U^{n+1}v - U^n v)$ tends to zero geometrically. Combining this with (2.1) and (2.2) we get the following

Theorem 1. If (2.4) holds for some $\gamma > 0$ then for some $g^* \in \mathbb{R}$, $v^* \in \mathbb{R}^N$

i) $\quad U^{n+1}v - U^n v = g^*.e + O((1-\gamma)^{n/N-1}) \quad (n \to \infty)$

ii) $\quad U^n v = ng^*.e + v^* + O((1-\gamma)^{n/N-1}) \quad (n \to \infty)$

iii) $\quad Uv^* = v^* + g.e .$

Proof. i) follows immediately from the geometric convergence of $sp(U^{n+1}v - U^n v)$ and (2.1) and (2.2) from which we have for all n and v:

$$\nabla(U^{n+1}v - U^n v) \leq \nabla(U^{n+2}v - U^{n+1}v) \leq \Delta(U^{n+2}v - U^{n+1}v) \leq \Delta(U^{n+1}v - U^n v) .$$

Now ii) follows from i) and iii) from ii). □

It is a direct consequence of theorem 1i) and lemma 1 that we will also find $C(1-\gamma)^{n/N-1}$-optimal stationary strategies in step n of a successive approximation procedure where C is some constant depending on the scrapvector v.

It remains to prove our main theorem

Theorem 2. If both the unichain assumption and assumption (1.2) are satisfied then there exists a $\gamma > 0$ such that for all Markov strategies π,ρ,π',ρ' and states $i,j \in S$

$$(2.5) \qquad \sum_{k\in S} \min\{\mathbb{P}_i^{\pi,\rho}(S_{N-1} = k), \mathbb{P}_j^{\pi,\rho}(S_{N-1} = k)\} \geq \gamma .$$

Proof. First we prove that the left hand side in (2.5) is positive for all $i,j \in S$ and pure Markov strategies.

We will sometimes use that π,ρ,π',ρ' may be denoted by $(f_0,f_1,\ldots)$, $(h_0,h_1,\ldots),(f_0',f_1',\ldots)$ and $(h_0',h_1',\ldots)$ respectively. Now fix the pure Markov strategies π,ρ,π',ρ' and defined $V_{N-1} := \{(s,s) \mid s \in S\}$ and $V_n := \{(t_1,t_2) \mid t_1,t_2 \in S$ and there exists a pair $(s_1,s_2) \in V_{n+1}$ such that

$$\min\{\mathbb{P}_{t_1}^{f_n,h_n}(S_1 = s_1),\mathbb{P}_{t_2}^{f_n',h_n'}(S_1 = s_2)\} > 0\}, \quad n = N-2,\ldots,0 .$$

Then V_0 becomes the set of all pairs (i,j) for which the left hand side of (2.5) is nonnegative. So what we have to show is $V_0 = S^2$. This may be done as follows. Define

$$V_n(s) := \{t \in S \mid (t,s) \in V_n\} .$$

Now we prove by induction, writing $|V_n(s)|$ for the number of elements in $V_n(s)$,

$$|V_n(s)| \geq N-n, \quad n = N-1,\ldots,0 . \tag{2.6}$$

First observe $V_{n+1} \subset V_n$ because of assumption (1.2) thus

$$|V_n(s)| \geq |V_{n+1}(s)| . \tag{2.7}$$

By definition we have $|V_{N-1}(s)| = |\{s\}| = 1$.
Now assuming (2.5) holds for $k = N-1,\ldots,n$ we prove it to hold for $n-1$.
We distinguish three cases

a) $|V_n(s)| > N-n$ then from (2.7) $|V_{n-1}(s)| \geq N-n+1$.
b) $|V_n(s)| = N-n$, $|V_{n-1}(s)| \geq N-n+1$.
c) $|V_n(s)| = |V_{n-1}(s)| = N-n$.

It is clear that in order to prove (2.5) it is sufficient to show that c) leads to a contradiction. Assume c).
We will write $\bar{V}$ for the complement of a subset V of S. Clearly for all $t \in \bar{V}_{n-1}(s)$, and all $(s_1,s_2) \in V_n$

$$\min\{\mathbb{P}_t^{f_{n-1},h_{n-1}}(S_1 = s_1),\mathbb{P}_s^{f_{n-1}',h_{n-1}'}(S_1 = s_2)\} = 0 \tag{2.8}$$

since otherwise $(t,s) \in V_{n-1}$ and thus $t \in V_{n-1}(s)$.

Specifying (2.8) for $(s_1,s_2) = (s_1,s)$ and using $\mathbb{P}_s^{f'_{n-1},h'_{n-1}}(S_1 = s) \geq \alpha > 0$ we see $\mathbb{P}_t^{f_{n-1},h_{n-1}}(S_1 = s_1) = 0$ for all $s_1 \in V_n(s)$.
This implies that under the pair of stationary strategies f_{n-1},h_{n-1} the set $\bar{V}_n(s) = \bar{V}_{n-1}(s)$ contains a recurrent subchain. Thus by the unichain assumption there must exist for any pair of pure policies f and h a state $s_* \in V_{n-1}(s)$ and a state $t \in \bar{V}_{n-1}(s)$ such that $\mathbb{P}_{s_*}^{f,h}(S_1 = t) > 0$. Otherwise it would be possible to construct from f'_{n-1},h'_{n-1} and f,h a third pair of policies with at least two different recurrent subchains. So we see that there exists a state $S_{N-2} \in V_{n-1}(s)$ and a state $t_{N-2} \in \bar{V}_{n-1}(s)$ such that $\mathbb{P}_{s_{N-2}}^{f'_{N-2},h'_{N-2}}(S_1 = t_{N-2}) > 0$, hence $(t_{N-2},s_{N-2}) \in V_{N-2}$.
But then again there must be a state $s_{N-3} \in V_{n-1}(s)\backslash\{s_{N-2}\}$ and a state $t \in \bar{V}_{n-1}(s) \cup \{s_{N-2}\}$ such that $\mathbb{P}_{s_{N-3}}^{f'_{N-3},h'_{N-3}}(S_1 = t) > 0$, and also a state $t_{N-3} \in \bar{V}_{n-1}(s)$ such that $(t_{N-3},s_{N-3}) \in V_{N-3}$. Continuing to reason in this way one shows that there are at least $N-n$ different elements $s_1 \in V_{n-1}(s)$ satisfying for some $t \in \bar{V}_{n-1}(s)$ $(t,s_1) \in V_{n-1}$. Hence, since $|V_{n-1}(s)| = N-n$ and $s \in V_{n-1}(s)$, also $(t,s) \in V_{n-1}$ for some $t \in \bar{V}_{n-1}(s)$. But this would imply $t \in V_{n-1}(s)$. Contradiction. So we conclude $|V_{n-1}(s)| \geq N-n+1$. Hence $V_0(s) = S$ for all $s \in S$ and therefore $V_0 = S^2$.
Since S, K_i, L_i are all finite there must exist a $\gamma > 0$ such that for all pure Markov strategies π,ρ,π',ρ'

$$(2.9) \qquad \min_{i,j} \sum_{k \in S} \min\{\mathbb{P}_i^{\pi,\rho}(S_{N-1} = k), \mathbb{P}_j^{\pi',\rho'}(S_{N-1} = k)\} \geq \gamma .$$

Moreover it is fairly obvious that the minimum of the left-hand side of (2.9) within the set of all Markov strategies equals the minimum within the set of pure Markov strategies π,ρ,π',ρ'. And this completes the proof. □

So we see that if the unichain assumption is satisfied and also 1.2 holds (for example as result of Schweitzer's transformation) then the method of successive approximations yields an ε-band for the value of the game as well as stationary ε-optimal strategies. And this exponentially fast.

3. The Functional Equation Uv = v + g.e has a Solution

In this section we assume that the functional equation Uv = v + g.e has a solution v^*, g^*. Moreover we assume throughout this section that (1.2) holds, i.e. $P(f,h) - \alpha I \geq 0$ for some $\alpha > 0$ and all f and h. We will show that in this case the method of successive approximations again yields an ε-band for the value g^* and stationary ε-optimal strategies.

First we show that if we take an arbitrary scrap vector $v \in \mathbb{R}^N$ then the sequences $\{U^n v - ng^*e\}$ and $\{U^{n+1}v - U^n v\}$ are bounded. Hence they both have convergent subsequences with limits, say $\tilde{v}$ and $\tilde{g}$. Then, essentially using (1.2), we show $\tilde{g} = g^*.e$. And finally we conclude that the above sequences are convergent.

Choose $v_0 \in \mathbb{R}^N$ arbitrary and define $v_{n+1} := Uv_n$ and $g_n := v_{n+1} - v_n$, $n = 0,1,\ldots$.
As in the proof of theorem 1 one easily argues

(3.1) $\quad \nabla g_n \leq \nabla g_{n+1} \leq \Delta g_{n+1} \leq \Delta g_n$.

As a consequence the sequence $\{g_n\}_{n\in\mathbb{N}}$ contains a convergent subsequence.

We also need the following inequalities formulated in

Lemma 3.

$$ng^*e + v^* + \nabla(v_0 - v^*).e \leq v_n \leq ng^*e + v^* + \Delta(v_0 - v^*).e \ ,$$

where g^*, v^* is the solution of (1.1).

Proof. $v_0 - v^* \geq \nabla(v_0 - v^*).e$, hence

$$v_n = U^n v_0 \geq U^n(v^* + \nabla(v_0 - v^*).e) = ng^*e + v^* + \nabla(v_0 - v^*).e \ .$$

Similarly one proves the second inequality. □

So also any subsequence of $\{v_n - ng^*e\}_{n\in\mathbb{N}}$ has a convergent subsequence. Now let $\tilde{v}$ be a limit point of the sequence $\{v_n - ng^*e\}_{n\in\mathbb{N}}$ and $\{v_{m_k} - m_k g^*e\}$ a subsequence converging to $\tilde{v}$. Let further $\tilde{g} \in \mathbb{R}^N$ be a limit-point of the sequence $\{g_{m_k}\}_{k\in\mathbb{N}}$ and $\{g_{n_\ell}\}_{\ell\in\mathbb{N}}$ a subsequence of $\{g_{m_k}\}$ converging to $\tilde{g}$. Then we have

Lemma 4. $U\tilde{v} = \tilde{v} + \tilde{g}$.

Proof. $U(v_{n_\ell} - n_\ell g^* e) = v_{n_\ell} - n_\ell g^* e + v_{n_\ell+1} - v_{n_\ell}$.

Taking the limit for $\ell \to \infty$ and using the finiteness of S, K_i and L_i, $i \in S$ to interchange sums and limits in $U(v_{n_\ell} - n_\ell g^* e)$ we get $U\tilde{v} = \tilde{v} + \tilde{g}$. □

Now we must show $\tilde{g} = g^*.e$. For this we need some additional lemmas.

Lemma 5. For all $\varepsilon > 0$ and all $M \in \mathbb{N}$ there exists an $n > M$ such that

$$\| U^n\tilde{v} - \tilde{v} - ng^* e\| \leq \varepsilon .$$

Proof. Let L be sufficiently large such that $\| v_{n_\ell} - n_\ell g^* e - \tilde{v}\| < \varepsilon/2$ for $\ell \geq L$. Then with $n = n_{\ell+k} - n_\ell$, $\ell \geq L$, k such that $n \geq M$

$$\| U^n\tilde{v} - \tilde{v} - ng^* e\| = \| U^n\tilde{v} - U^n(v_{n_\ell} - n_\ell g^* e) + v_{n_{\ell+k}} - \tilde{v} - n_{\ell+k} g^* e\|$$

$$\leq \| U^n\tilde{v} - U^n(v_{n_\ell} - n_\ell g^* e)\| + \| v_{n_{\ell+k}} - \tilde{v} - n_{\ell+k} g^* e\| < \varepsilon$$

where we used $\| U^n v - U^n w\| \leq \| v - w\|$ for all n,v,w. □

Lemma 6. For all $\varepsilon > 0$ and for all $M \in \mathbb{N}$ there exists an $n > M$ such that

$$\| U^{n+1}\tilde{v} - U^n\tilde{v} - \tilde{g}\| < 2\varepsilon .$$

Proof. Choose n as in lemma 5 then

$$\| U^{n+1}\tilde{v} - U^n\tilde{v} - \tilde{g}\| = \| U(U^n\tilde{v}) - U(\tilde{v} - ng^* e) + U\tilde{v} - \tilde{v} - \tilde{g} + \tilde{v} + ng^* e - U^n\tilde{v}\| \leq 2\varepsilon .□$$

Lemma 7. $\Delta(U^{k+1}\tilde{v} - U^k\tilde{v}) = \Delta\tilde{g}$ for all $k = 0,1,\ldots$.

Proof. Obviously $\Delta(U^{k+1}\tilde{v} - U^k\tilde{v}) \leq \Delta(U\tilde{v} - \tilde{v}) = \Delta\tilde{g}$. So it remains to show

$$\Delta(U^{k+1}\tilde{v} - U^k\tilde{v}) \geq \Delta\tilde{g} .$$

From lemma 6 we know that for all $\varepsilon > 0$ there exists an $n > k$ such that

$$\Delta(U^{n+1}\tilde{v} - U^{n}\tilde{v}) > \Delta\tilde{g} - \varepsilon ,$$

hence with

$$\Delta(U^{n+1}\tilde{v} - U^{n}\tilde{v}) \leq \Delta(U^{k+1}\tilde{v} - U^{k}\tilde{v}) ,$$

we have

$$\Delta(U^{k+1}\tilde{v} - U^{k}\tilde{v}) > \Delta\tilde{g} - \varepsilon$$

for all $\varepsilon > 0$ so

$$\Delta(U^{k+1}\tilde{v} - U^{k}\tilde{v}) \geq \Delta\tilde{g}$$

which completes the proof. □

This lemma is especially important since we may combine it with lemmas 8 and 9 below.

Define $\Delta I_n := \{i \mid g_n(i) = \Delta g_n\}$.

Lemma 8. If $\Delta g_{n+1} = \Delta g_n$ then $\Delta I_{n+1} \subset \Delta I_n$.

Proof. Let f_{n+1} and h_n satisfy $L(f_{n+1},h)v_n \geq v_{n+1}$ and $L(f,h_n)v_{n-1} \leq v_n$ for all f and h. Since $P(f_{n+1},h_n) = (1-\alpha)Q(f_{n+1},h_n) + \alpha I$, where $Q(f_{n+1},h_n)$ is a stochastic matrix as well, we have (cf. 2.1).

$$g_{n+1} \leq P(f_{n+1},h_n)g_n = [(1-\alpha)Q(f_{n+1},h_n) + \alpha I]g_n ,$$

hence for $i \in \Delta I_{n+1}$

$$\Delta g_{n+1} = g_{n+1}(i) \leq (1-\alpha)Q(f_{n+1},h_n)\Delta g_n.e + \alpha g_n(i) = (1-\alpha)\Delta g_n + \alpha g_n(i).$$

So with $\Delta g_{n+1} = \Delta g_n$ we get $g_n(i) = \Delta g_n$, which implies $i \in \Delta I_n$. □

Lemma 9. If $\Delta g_n = \Delta g_N$ for all $n \geq N$ then $\Delta g_N = g^*$.

Proof. From $\Delta I_{n+1} \subset \Delta I_n$ and $\Delta I_n \neq \emptyset$ we see that there must exist an $i \in \bigcap_{n\geq N} \Delta I_n$. For this i we have

$$v_{N+k}(i) = v_{N+k}(i) - v_{N+k-1}(i) + \ldots + v_{N+1}(i) - v_N(i) + v_N(i)$$

$$= n\Delta g_N + v_n(i) .$$

With lemma 3 this implies $\Delta g_N = g^*$. □

Now we are able to prove the main result of this section.

Theorem 3. $\tilde{g} = g^*.e$.

Proof. Define $v_0 := \tilde{v}$. Then we obtain with lemma 6 and the reasoning of lemmas 3 and 9 $\Delta\tilde{g} = g^*$. Adapting lemma's 9 and 7 one may also show $\nabla\tilde{g} = g^*$. Hence $\tilde{g} = g^*.e$. □

Since the limitpoint $\tilde{g}$ is a constant vector, $\tilde{g}$ is also the limit of the sequence g_n (because of the monotonicity of Δg_n and ∇g_n).
And also $\tilde{v}$ is the limit of the sequence $\{v_n - ng^*e\}$ as follows from

Theorem 4. $v_n = \tilde{v} + ng^*e + o(1)$ $(n \to \infty)$.

Proof. Obviously since $\tilde{g} = g^*e$ $U^p\tilde{v} - \tilde{v} = pg^*e$ for all $p \in \mathbb{N}$. Now fix $\varepsilon > 0$. Let k satisfy $\| v_{n_k} - n_k g^*e - \tilde{v}\| < \varepsilon$ then for all $n > n_k$, writing $p = n - n_k$.

$$\| v_n - ng^*e - \tilde{v}\| = \| U^p(v_{n_k} - n_k g^*e) - U^p\tilde{v} + U^p\tilde{v} - \tilde{v} - pg^*e\|$$

$$\leq \| v_{n_k} - n_k g^*e - \tilde{v}\| + \| U^p\tilde{v} - \tilde{v} - pg^*e\| < \varepsilon .$$

Hence for all $\varepsilon > 0$ there exists a K such that for all $n > K$ $\| v_n - ng^*e - \tilde{v}\| < \varepsilon$.
This completes our proof. □

So we see that if the functional equation 1.1 has a solution then the method of successive approximations yields an ε-band for this value and stationary ε-optimal strategies for both players.
Probably the convergence is again exponentially fast but we have not proved this yet.

References

[1] D. Blackwell and T.S. Ferguson, The big match, Ann. Math. Statist. 39 (1968), pp. 159-163.

[2] C. Derman and R. Strauch, A note on memoryless rules for controlling sequential control processes, Ann. Math. Statist. 37 (1966), pp. 276-278.

[3] A. Federgruen, On N-person stochastic games with denumerable state space, Mathematical Centre report BW 67/76, Mathematical Centre, Amsterdam, 1976.

[4] D. Gillette, Stochastic games with zero stop probabilities, Contributions to the theory of games, Vol. III, M. Dresher, A.W. Tucker, P. Wolfe, eds., Princeton University Press, Princeton, New Jersey, pp. 179-187.

[5] A. Hoffman and R. Karp, On non-terminating stochastic games, Management Science, 12 (1966), pp. 359-370.

[6] P.D. Rogers, Non-zero sum stochastic games, Ph. D. thesis submitted to the University of California, Berkeley, 1969.

[7] P.J. Schweitzer, Iterative solution of the functional equations of undiscounted Markov renewal programming, J. Math. Anal. Appl. 34 (1971), pp. 495-501.

[8] L.S. Shapley, Stochastic games, Proc. Nat. Acad. Sci. USA, 39 (1953), pp. 1095-1100.

[9] M. Sobel, Noncooperative stochastic games, Ann. Math. Statist. 42 (1971), pp. 1930-1935.

[10] M.A. Stern, On stochastic games with limiting average payoff, Ph. D. thesis submitted to the University of Illinois, Chicago, 1975.

[11] L.E. Zachrisson, Markov games, Advances in game theory, M. Dresher, L.S. Shapley, A.W. Tucker, eds., Princeton University Press, Princeton, New Jersey, pp. 211-253, 1964.