stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE
(DEPARTMENT OF OPERATIONS RESEARCH)

BW 102/79 MAART

O.J. VRIEZE

CHARACTERIZATION OF OPTIMAL STATIONARY STRATEGIES IN UNDISCOUNTED STOCHASTIC GAMES

Preprint

2e boerhaavestraat 49 amsterdam

BIBLIOTHEEK MATHEMATISCH CENTRUM
AMSTERDAM

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): 90D15, 90D05, 90C40

Characterization of optimal stationary strategies in undiscounted stochastic games *)

by

O.J. Vrieze

ABSTRACT

This paper considers two-person zero-sum undiscounted stochastic games, with finite state space and finite action spaces for the both players. The set of games having as well a value as optimal stationary strategies for the both players is characterized, by using a set of three optimality equations. Next the sets of optimal stationary strategies are characterized.

Furthermore the set of games, for which the value is independent of the initial state is analysed in detail.

KEY WORDS & PHRASES: *stochastic games, average payoff optimality, initial state independent value, optimal stationary strategies*

---

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

This paper deals with two-person zero-sum undiscounted stochastic games. As well the state space as the sets of pure actions of the both players are assumed to be finite sets. It is not yet known, whether all games of this type have a value. BEWLEY and KOHLBERG ([1] and [2]) suggested a candidate for being the value of such a game, if this value exists. Until now no game is constructed with a value unequal to this candidate. They also determined a certain subclass of the games, which have a value and for which furthermore both players have optimal stationary strategies. They showed that their beautiful attack circumvent all earlier results on this topic.

In this paper (section 5) we will characterize the set of all games, which have a value and for which both players have optimal stationary strategies. In deducing these results we use a set of three optimality equations, which may be seen as the natural extensions of the set of two optimality equations, with which Markov decision problems are solved (FEDERGRUEN [5], page 44; see also section 3 of this paper). The existence of a solution of these three optimality equations proves to be equivalent to the existence of the value of the game together with the existence of optimal stationary strategies for both players.

In [5] FEDERGRUEN stated a set of two optimality equations, showing that the existence of a solution of these equations is a necessary condition for the existence of optimal stationary strategies for the both players. However he himself gave an example (page 174), which obstructed a sufficiency theorem.

In section 4 we in detail consider the set of games, for which the candidate of being the value is state-independent. We will show, that these games have a value and furthermore that both players have optimal (Markov-) strategies and also that both players have at least ε-optimal stationary strategies.

In section 3 we state some well-known results, concerning Markov decision problems and in section 2 the necessary preliminaries are given.

2. PRELIMINARIES

A two-person zero-sum stochastic game, notated Γ, is characterized by a five-tuple $\langle S, \{A_1(k);\ k \in S\}, \{A_2(k);\ k \in S\}, r, P\rangle$, where S and $A_i(k)$, $i \in \{1,2\}$, $k \in S$ are non-empty sets and r and P are mappings. When S and each $A_i(k)$, $i \in \{1,2\}$, $k \in S$, are finite sets, then we will call such a game a finite stochastic game and it is this type of game, which will be considered in this paper.

The parameters of Γ have the following meaning:

- $S = \{1,2,\ldots,N\}$ is called the state space.
- $A_i(k) = \{1,2,\ldots,j_{ik}\}$, $i \in \{1,2\}$, $k \in S$ is called the set of pure actions for player i in state k.
- r is a real valued function defined on the set of triples $T = \{(k,a_1,a_2);\ k \in S,\ a_1 \in A_1(k),\ a_2 \in A_2(k)\}$ and is called the payoff function.
- P is a map from T into the set $\mathcal{P}(S)$ of probability measures on S and is called the transition probability map.

Such a stochastic game corresponds with a dynamic system, where the dynamic behaviour as well as the rewards are influenced by the players at discrete points in time (called stages), say $t = 0,1,2,\ldots$, in the following way:

At each stage t the players observe the current state of the system. They, then, have to select, independently of one another, an action. If at time t the system is in state $k \in S$ and if player 1 selects action $a_1 \in A_1(k)$ and player 2 action $a_2 \in A_2(k)$, then two things happen.

(1) player 1 obtains an immediate reward $r(k,a_1,a_2)$ from player 2.
(2) the system moves with probability $P(k,a_1,a_2)\{\ell\}$ - which we denote by $p(\ell|k,a_1,a_2)$ from now on - to state $\ell \in S$, which will be observed at the next stage t+1.

For a finite set $B = \{1,2,\ldots,j\}$ we denote by $\mathcal{P}(B)$ the set of probability measures on B. Note that there is a one-to-one correspondence between the set $\mathcal{P}(B)$ and the set $\{x;\ x = (x_1,x_2,\ldots,x_j),\ x_n \geq 0,\ n \in B,\ \sum_{n=1}^{j} x_n = 1\}$ and this last representation is the one, which will be used in the sequel.

A history dependent strategy for player i is a rule, which, for each

stage $t \in \{0,1,2,\ldots\}$ and each state $k \in S$, selects dependent of t and the history of the game at time t an element of $P(A_i(k))$, according to which player i should choose his pure action in that situation, if he adopts that strategy. The history of the game at time t is the sequence of states and actions, that actually have occured until time t. The set of history dependent strategies for player i is notated as H_i.

A Markov strategy for player i is a history dependent strategy, such that for each stage $t \in \{0,1,2,\ldots\}$ and each state $k \in S$ the seclection of an element of $P(A_i(k))$ is independent of the history of the game at time t.

A stationary strategy for player i is a Markov strategy, such that for each stage $t \in \{0,1,2,\ldots\}$ and each state $k \in S$ the selection of an element of $P(A_i(k))$ is independent of t. The set of stationary strategies for player i will be notated as ST_i.

A strategy for player i will be notated as π_i; if π_i is a stationary strategy, then π_i can be notated as $\pi_i = (\pi_{i1},\ldots,\pi_{iN})$, where $\pi_{ik} \in P(A_i(k))$, $k \in S$.

If the players 1 and 2 play strategy π_1 respectively strategy π_2, then a probability measure on the set of the infinite streams of payoffs is defined. Comparing of these probability measures can be done on different ways, determining different types of games. For each of these types in a certain way a N-vector is added to each probability measure, associated with a pair (π_1,π_2); the k'th component of this vector denotes the payoff, when π_1 and π_2 are played, for the specific game with state k as initial state.

Let for a certain type of games $W(\pi_1,\pi_2)$ be the vector associated to the pair of strategies (π_1,π_2).

Player 1 wants to maximize $W(\pi_1,\pi_2)$ and player 2 wants to minimize this vector, both componentwise.

On each type of games the following definitions can be applied.

DEFINITION 1. A game is said to have a *value* if componentwise

$$\inf_{\pi_2 \in H_2} \sup_{\pi_1 \in H_1} W(\pi_1,\pi_2) = \sup_{\pi_1 \in H_1} \inf_{\pi_2 \in H_2} W(\pi_1,\pi_2).$$

DEFINITION 2. If a player has a value, say W, then for $\varepsilon \geq 0$ a strategy π_1 for player 1 is called ε-*optimal* if componentwise

$$\inf_{\pi_2 \in H_2} W(\pi_1, \pi_2) \geq W - \varepsilon.$$

A strategy π_2 for player 2 is called ε-optimal if componentwise

$$\sup_{\pi_1 \in H_1} W(\pi_1, \pi_2) \leq W + \varepsilon$$

0-optimal strategies are usually called optimal strategies.

From now on $\inf_{\pi_2 \in H_2}$ and $\sup_{\pi_1 \in H_1}$ will be abbreviated to $\inf_{\pi_2}$ and $\sup_{\pi_1}$ and as inequalities between vectors will always be componentwise we shall omit the specification "componentwise".

DEFINITION 3. A t-*step stochastic game* is a stochastic game which will stop after t moves of the game, possbily with a terminating state dependent terminal payoff; in such games the expected payoff per stage are merely added up.

Each two-person zero-sum finite t-step stochastic game has a value (SHAPLEY [7]) and both players have optimal Markov-strategies. In the following $V_t = (V_t(1),\ldots,V_t(N))$ will denote the value of a t-step stochastic game without a terminal payoff (or equivalently with terminal payoff 0).

DEFINITION 4. A β-*discounted stochastic game* is a stochastic game with infinite stages, in which a payoff r_t at time t will be discounted by a factor β^t, with $\beta \in [0,1)$, which results in an evaluation of $\beta^t r_t$ at time 0 of this payoff. For a pair of strategies (π_1,π_2) the discounted expected payoffs are added up.

Each two-person zero-sum finite β-discounted stochastic game has a value and both players have optimal stationary strategies (SHAPLEY [7]). In the following $V_\beta = (V_\beta(1),\ldots,V_\beta(N))$ will denote the value of the β-discounted stochastic game.

DEFINITION 5. An *undiscounted stochastic game* is a stochastic game with infinite stages, in which the evaluation of the stream of expected payoffs is carried out by computing the expected average payoff per step.

However this average need not exist and this gives rise to more detailed definitions. In BEWLEY and KOHLBERG [2] six ways of computing an expected average payoff are mentioned. They prove the following lemma:

LEMMA 2.1. *If for a two-person zero-sum finite undiscounted stochastic game a stationary strategy guaranteed a player an amount* V *for one of the six ways of computing an expected average payoff, then this stationary strategy guarantees this amount* V *in all six ways of computing an average payoff.*

As in this paper we will consider games, in which both players have ε-optimal or optimal stationary strategies, we may choose arbitrary one of these six ways of computing the expected average payoff. We will choose the so-called limit expected average criterion, where the average payoff for a pair of strategies (π_1,π_2), notation $V(\pi_1,\pi_2)$, is computed as

$$V(\pi_1,\pi_2) = \liminf_{T\to\infty} E_{\pi_1\pi_2}\{\frac{1}{T+1}\sum_{t=0}^{T} \underline{r}_t(\pi_1,\pi_2)\},$$

where $\underline{r}_t(\pi_1,\pi_2)$ is the random variable determined by π_1 and π_2 and denoting the payoff at time t and where $E_{\pi_1\pi_2}$ means the expectation with respect to π_1 and π_2 of the expression between the accolades.

Let for a pair of stationary strategies (π_1,π_2) the quantities $p(\ell|k,\pi_1,\pi_2)$ and $r(\pi_1,\pi_2) = (r(1,\pi_1,\pi_2),\ldots,r(N,\pi_1,\pi_2))$ be defined as:

$$P(\ell|k,\pi_1,\pi_2) = \sum_{a_1=1}^{j_{1k}} \sum_{a_2=1}^{j_{2k}} \pi_{1k}(a_1).\pi_{2k}(a_2).p(\ell|k,a_1,a_2)$$

and

$$r(k,\pi_1,\pi_2) = \sum_{a_1=1}^{j_{1k}} \sum_{a_2=1}^{j_{2k}} \pi_{1k}(a_1).\pi_{2k}(a_2).r(k,a_1,a_2).$$

With a pair of stationary strategies (π_1,π_2) we can associate a matrix $P_{\pi_1\pi_2}$, with the (k,ℓ)-element equal to $p(\ell|k,\pi_1,\pi_2)$, $k \in \{1,\ldots,N\}$, $\ell \in \{1,\ldots,N\}$.

Let

$$Q_{\pi_1\pi_2} = \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} P^t_{\pi_1\pi_2}$$

be the Cesaro-limit of $P_{\pi_1\pi_2}$, where $P^0_{\pi_1\pi_2}$ is the unit matrix and

$$P^t_{\pi_1\pi_2} = P_{\pi_1\pi_2}(P^{t-1}_{\pi_1\pi_2}), \quad t \in \{1,2,\ldots\}.$$

It can be seen that for a pair of stationary strategies the expected average payoff per stage equals

$$V(\pi_1,\pi_2) = \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} P^t_{\pi_1\pi_2}.r(\pi_1,\pi_2) = Q_{\pi_1\pi_2}.r(\pi_1,\pi_2).$$

Let for a pair of stationary strategies (π_1,π_2) and associated transition probability matrix $P_{\pi_1\pi_2}$ the set $R_{\pi_1\pi_2}$ be the set of states which are recurrent under π_1 and π_2.

A matrix game with pure action sets A_1 and A_2 and payoff function $r: A_1 \times A_2 \to \mathbb{R}$ will be notated as $\langle r \rangle$, while $\mathrm{Val}(\langle r \rangle)$ stands for the value of this matrix game.

Let $v = (v(1),\ldots,v(N)) \in \mathbb{R}^N$. For a stochastic game, let $P(v) = (P_1(v),\ldots,P_N(v))$ denote the vector with as k'th component the real-valued function $P_k(v): A_1(k)\times A_2(k) \to \mathbb{R}$, where

$$P_k(v)(a_1,a_2) = \sum_{\ell=1}^{N} p(\ell|k,a_1,a_2).v(\ell).$$

$$\mathrm{Val}(\langle r+P(v)\rangle) = (\mathrm{Val}_1(\langle r_1+P_1(v)\rangle),\ldots,\mathrm{Val}_N(\langle r_N+P_N(v)\rangle))$$

will denote the vector with as k'th component the value of the matrix game $\langle r_k+P_k(v)\rangle$, where $r_k+P_k(v): A_1(k)\times A_2(k) \to \mathbb{R}$ is defined as

$$(r_k+P_k(v))(a_1,a_2) = r(k,a_1,a_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,a_2).v(\ell).$$

Two important results of BEWLEY and KOHLBERG [1] are stated in the following lemma's:

LEMMA 2.2. *For each two-person zero-sum finite stochastic game there exist a vector* $g^* = (g^*(1),\ldots,g^*(N))$ *such that:*

$$\lim_{t\to\infty} \frac{V_t}{t} = \lim_{\beta\uparrow 1} (1-\beta)V_\beta = g^*$$

and g^* *obeys*

$$g^* = \mathrm{Val}(<P(g^*)>).$$

For the rest of this paper g^* always stands for the limit mentioned in lemma 2.2.

LEMMA 2.3. *Let* $\rho = \frac{1-\beta}{\beta}$, *then for a two-person zero-sum finite stochastic game, we can expand* V_β *(also notated* V_ρ*) for* β *close enough to* 1 *(*ρ *close enough to* 0*) in a Puisseux series of the following form:*

$$V_\rho = \sum_{k=-M}^{\infty} a_k \cdot \rho^{k/M},$$

where M *is a positive integer and* a_k *a vector for each* k. *Furthermore it holds that* $a_{-M} = g^*$.

A further nice statement in BEWLEY and KOHLBERG [2] is the fact, that neither player can, using stationary strategies, guarantee himself more than g^*. This leads to the following lemma.

LEMMA 2.4. *If a two-person zero-sum finite undiscounted stochastic game has a value and if both players have* ε*-optimal stationary strategies, then the value equals* g^*.

It may be clear, that also in general g^* is a high candidate for being the value of an undiscounted stochastic game, however this still remains an open problem.

3. MARKOV DECISION PROBLEMS

If in each state a player, say player 1, has only one action, then he is called a dummy player (he has no real choice). Then the game is identical

to a Markov decision problem, in which the total payoff should be minimized. Markov decision problems are analysed in detail, e.g. DENARDO and FOX [3], DERMAN [4] and FEDERGRUEN [5].

It is known that there always exist optimal stationary strategies. The following lemma can be found in DENARDO and FOX [3]: (Notations are analogous to the one introduced in section 2).

LEMMA 3.1. *For a minimizing finite Markov decision problem, there exist a unique vector* g *and a non-unique vector* v *such that*

$$(1) \qquad g = \min_{\pi_2 \in ST_2} (<P(g)>)$$

$$(2) \qquad v+g = \min_{\pi_2 \in O} (<r+P(v)>)$$

where $O = \sum_{k=1}^{N} O_k$ *and*

$$O_k = \{\pi_{2k} \mid g(k) = \sum_{a_2=1}^{j_{2k}} \pi_{2k}(a_2) \sum_{\ell=1}^{N} p(\ell|k,a_2)g(\ell)\}.$$

Then g *is the minimal expected average payoff and a stationary strategy* π_2 *is optimal if and only if* (a) $\pi_2 \in O$ *and* (b) *for each* $k \in R_{\pi_2}$ *it holds that* π_{2k} *is optimal in* (2).

Let $R^* = \{k \mid$ there exist an optimal stationary strategy π_2 with $k \in R_{\pi_2}\}$ FEDERGRUEN [5] proved, that there exist an optimal stationary strategy π_2, such that $R^* = R_{\pi_2}$.

If player 2 is a dummy player, then we have a mazimizing Markov decision problem and for the solution of this problem min in lemma 3.1 should be replaced by max.

Now suppose for a two-person zero-sum finite stochastic game Γ, that player 1 plays the stationary strategy π_1. We will ask for the best answer of player 2 to this strategy π_1.

Consider therefore the following minimizing Markov decision problem Γ': set of states: $\{1,2,\ldots,N\}$; set of pure actions in state k: $A_2(k)$; rewards r'_k: $A_2(k) \to \mathbb{R}$ defined as $r'_k(a_2) = r(k,\pi_1,a_2)$; transition probabilities defined as $p'(\ell|k,a_2) = p(\ell|k,\pi_1,a_2)$. Now it can be seen, that if a strategy

for player 2 in the original game Γ is such, that for each time t, his selection of his action does not depend on the actual past actions of player 1, then this strategy can also be interpretated as a strategy in the derivated Markov decision problem Γ' and moreover the expected average payoff in both problems are the same. This especially holds for the class of stationary strategies. Along the same lines of theorem 1, page 91 in DERMAN [4] it can be shown, that if player 1 uses a strategy π_2 in Γ which does depend on the past actions of player 1, then there can be constructed a strategy $\tilde{\pi}_2$, not depending on the past actions of player 1, such that for each t the expected payoff at time t remains equal.

Summarizing the above yields, that for looking at the best answer of player 2 to a stationary strategy π_1 of player 1 we may consider the above stated Markov decision problem Γ'. So by lemma 3.1 there exist a stationary strategy, which is a best answer to π_1.

This will be stated in a lemma.

LEMMA 3.2. *If for a two-person zero-sum finite stochastic game* π_1 *is a stationary strategy for player* 1, *then*

$$\inf_{\pi_2 \in H_2} V(\pi_1,\pi_2) = \min_{\pi_2 \in ST_2} V(\pi_1,\pi_2).$$

The set R^* of the derivated Markov decision problem Γ' by a stationary strategy π_1, will be notated as $R^*(\pi_1)$.

Evidently, if we fix a stationary strategy π_2 for player 2, then the only change in the above analysis is that min should be replaced by max.

4. UNDISCOUNTED STOCHASTIC GAMES IN WHICH $g^* = \tilde{g}.\bar{1}$

In this section we will consider games in which g^* does not depend on the initial state. It appeared that these games have a value and that both players have optimal Markov strategies and ε-optimal stationary strategies. The next lemma will be needed:

LEMMA 4.1. *If* $\langle r_1 \rangle$ *and* $\langle r_2 \rangle$ *are matrix games of the same dimensions, then*

$$|\mathrm{Val}\langle r_1\rangle - \mathrm{Val}\langle r_2\rangle| \leq \max_{i,j} |r_1(i,j) - r_2(i,j)|.$$

Let $\|.\|$ *denote the max norm in* $\mathbb{R}^N$.

THEOREM 4.2. *For a two-person zero-sum finite undiscounted stochastic game the following two assertions are equivalent:*

(i) $g^* = \tilde{g}.\bar{1}$

(ii) $\exists \tilde{g} \in \mathbb{R}$ *such that* $\forall \varepsilon > 0$, $\exists v_\varepsilon \in \mathbb{R}^N$ *with*
$\|v_\varepsilon + \tilde{g}.\bar{1} - \mathrm{Val}(\langle r+P(v_\varepsilon)\rangle)\| \leq \varepsilon$.

PROOF. Suppose (i) is true. From discounted stochastic game theory we know $V_\beta = \mathrm{Val}(\langle r+\beta P(V_\beta)\rangle)$ (SHAPLEY [7]). Then for β close enough to 1 and using lemma 2.1 and lemma 4.1 we get:

$$\|V_\beta + \tilde{g}.\bar{1} - \mathrm{Val}(\langle r+P(V_\beta)\rangle)\| =$$

$$= \|\mathrm{Val}(\langle r+\beta P(V_\beta)+\tilde{g}.\bar{1}\rangle) - \mathrm{Val}(\langle r+P(V_\beta)\rangle)\| \leq$$

$$\leq \max_k \max_{i,j} |\tilde{g} + \beta P_k(V_\beta) - P_k(V_\beta)| =$$

$$= \max_k \max_{i,j} |\tilde{g} - (1-\beta) P_k(V_\beta)|$$

$$\leq \|\tilde{g}.\bar{1} - (1-\beta)V_\beta\| \leq \varepsilon.$$

So for β close enough to 1, we see that V_β obeys (2).

Suppose now that (ii) is true. Let π^ε_{1k} be an optimal action for player 1 in the matrix game $\langle r_k+P_k(v_\varepsilon)\rangle$ and consider the stationary strategy $\pi^\varepsilon_1 = (\pi^\varepsilon_{11},\ldots,\pi^\varepsilon_{1N})$. Then for all stationary strategies π_2 for player 2 we have

$$v_\varepsilon + \tilde{g}.\bar{1} - r(\pi^\varepsilon_1,\pi_2) - P_{\pi^\varepsilon_1\pi_2}(v_\varepsilon) \leq \varepsilon.\bar{1}.$$

Multiplying this inequality by $Q_{\pi^\varepsilon_1\pi_2}$ and rearranging terms yields

$$\min_{\pi_2 \in ST_2} Q_{\pi^\varepsilon_1\pi_2}.r(\pi^\varepsilon_1,\pi_2) \geq (\tilde{g}-\varepsilon).\bar{1}. \qquad (1)$$

Analogously we can show the existence of a stationary strategy π_2^ε for player 1, such that

$$\max_{\pi_1 \in ST_1} Q_{\pi_1 \pi_2^\varepsilon}.r(\pi_1,\pi_2^\varepsilon) \leq (\tilde{g}+\varepsilon).\bar{1}. \tag{2}$$

Clearly (1) and (2) together with lemma 3.2 show that $\tilde{g}.\bar{1}$ is the value of the game and as both players have ε-optimal stationary strategies we may conclude from lemma 2.3 that $\tilde{g}.\bar{1} = g^*$. □

COROLLARY 4.3. *If* (i) *or* (ii) *of theorem* 4.2 *holds, then*

(a) *the game has a value*

(b) *both players have ε-optimal stationary strategies*

(c) *both players have optimal Markov strategies.*

PROOF. (a) and (b) are shown in the second part of the proof of theorem 4.2. Concerning (c) let for $\delta > 0$ $t(\delta)$ be the smallest integer such that $\frac{V_t}{t} \geq (\tilde{g}-\delta).\bar{1}$, $t \geq t(\delta)$. Let π_1^δ be an optimal Markov strategy for player 1 in the $t(\delta)$-step game.

Consider with $0 < \delta < 1$ the following Markov strategy for player 1: in the first $t(\delta)$ steps he should play π_1^δ, in the next $t(\delta_3^2)$ steps he should play $\pi_1^{\delta^2}$; in the next $t(\delta^3)$ steps he should play $\pi_1^{\delta^3}$, etc. Then it can be seen, that for each initial state player 1, playing the above Markov strategy, guarantees himself:

$$\lim_{T\to\infty} \frac{\sum_{k=1}^{T} t(\delta^k)(\tilde{g}-\delta^k)}{\sum_{k=1}^{T} t(\delta^k)} = \tilde{g} - \lim_{T\to\infty} \frac{\sum_{k=1}^{T} t(\delta^k).\delta^k}{\sum_{k=1}^{T} t(\delta^k)} = \tilde{g}. \qquad \square$$

The following theorem characterizes for which games with $g^* = \tilde{g}.\bar{1}$ both players have optimal stationary strategies. This theorem can also be found in FEDERGRUEN [5].

THEOREM 4.4. *For a two-person zero-sum finite undiscounted stochastic game the following three assertions are equivalent:*

(i) $g^* = \tilde{g}.\bar{1}$ *and both players have optimal stationary strategies.*

(ii) $\exists v \in \mathbb{R}^N$ *and* $\tilde{g} \in \mathbb{R}$ *such that* $v + \tilde{g}.\bar{1} = \mathrm{Val}(<r+P(v)>)$.

(iii) $a_{-M+k} = 0$, $k \in \{1,2,\ldots,M-1\}$ *(coefficients in Puisseux series expansion of* V_ρ*) and* $g^* = \tilde{g}.\bar{1}$ *is the value of the game.*

PROOF. The proof runs as follows: (ii) ⇒ (i) ⇒ (iii) ⇒ (ii).

(ii) ⇒ (i) follows similarly as the proof of the second part of theorem 4.2.

(i) ⇒ (iii) this part needs some more knowledge of the Puisseux series expansion; the proof of it will be omitted but can be found in BEWLEY and KOHLBERG [2] (lemma 7.11).

(iii) ⇒ (ii) from the Puisseux series expansion one can derive

$$\lim_{\beta\uparrow 1} (V_\beta - \frac{\tilde{g}.\bar{1}}{1-\beta}) = a_0$$

Now

$$V_\beta = \mathrm{Val}(<r+\beta P(V_\beta)>) = \mathrm{Val}(<r+\beta P(V_\beta - \frac{\tilde{g}.\bar{1}}{1-\beta}) - \tilde{g}.\bar{1} + \frac{\tilde{g}.\bar{1}}{1-\beta}>)$$

or

$$V_\beta - \frac{\tilde{g}.\bar{1}}{1-\beta} + \tilde{g}.\bar{1} = \mathrm{Val}(<r+\beta P(V_\beta - \frac{\tilde{g}.\bar{1}}{1-\beta})>).$$

Then taking the limit for $\beta\uparrow 1$ from the left and right part of this expression yields $a_0 + \tilde{g}.\bar{1} = \mathrm{Val}(<r+P(a_0)>)$ and so (ii) is true. □

REMARK 4.5. From theorem 4.2 we see that a fourth equivalent assertion is: $\bigcap_{\varepsilon>0} V_\varepsilon$ is not empty, where for $\varepsilon > 0$ V_ε is defined as $V_\varepsilon = \{v_\varepsilon \mid v_\varepsilon \in \mathbb{R}^N$ and v_ε obeys part (ii) of theorem 4.2$\}$.

There are also games in which $g^* = \tilde{g}.\bar{1}$ and where only one of the players has optimal stationary strategies. In the same way as above it can be shown that in that case the following theorem holds (of course a similar theorem can be stated concerning player 2).

THEOREM 4.6. *For a two-person zero-sum finite undiscounted stochastic game the following two assertions are equivalent.*

(i) $g^* = \tilde{g}.\bar{1}$ *and only player* 1 *has an optimal stationary strategy.*

(ii) (a) $\exists \tilde{g} \in \mathbb{R}$ *such that* $\forall \varepsilon > 0$, $\exists v_\varepsilon \in \mathbb{R}^N$ *with*

$$\| v_\varepsilon + \tilde{g}.\bar{1} - \mathrm{Val}(<r+P(v_\varepsilon)>) \| \leq \varepsilon$$

(b) $\exists v \in \mathbb{R}^N$ *such that* $v+g.\bar{1} \leq \mathrm{Val}(<r+P(v)>)$, *where for all* v, *obeying this inequality, at least in one component this inequality is strict.*

5. CHARACTERIZATION OF UNDISCOUNTED STOCHASTIC GAMES HAVING OPTIMAL STATIONARY STRATEGIES

In this section we will characterize the set of undiscounted stochastic games, for which both players have optimal stationary strategies. Next we will characterize these sets of optimal stationary strategies.

We will need the following result concerning Markov decision problems. This lemma holds as well for a minimizing problem as for a maximizing problem.

LEMMA 5.1. *Let for an undiscounted Markov decision problem* Γ, *with optimal payoff* g, *the Markov decision problem* Γ' *equal to* Γ *except that the payoff function* r' *is defined as* $r' = r-g$, *i.e. if in state* k *the action* a *is taken, then the immediate payoff equals* $r'(k,a) = r(k,a)-g(k)$, *then the optimal payoff for the undiscounted Markov decision problem* Γ' *equals* $\bar{0}$.

PROOF. We will proof the lemma for a minimizing problem. From lemma 3.1 we derive $P_{\pi_2}.g \geq g$, $\forall \pi_2 \in ST_2$ and by multiplying this inequality by Q_{π_2} we see that $(P_{\pi_2}.g)(k) = g(k)$ for $k \in R_{\pi_2}$ (the recurrent states under π_2). This implies that $(Q_{\pi_2}.g)(k) = g(k)$ for $k \in R_{\pi_2}$. Then for arbitrary stationary strategy π_2 we get for $k \in R_{\pi_2}$:

$$(Q_{\pi_2}.(r(\pi_2)-g))(k) = (Q_{\pi_2}.r(\pi_2))(k)-g(k) \geq g(k)-g(k) = 0.$$

But as the average payoff in the transient states is a convex combination of the average payoffs in the recurrent classes it follows that

$$\min_{\pi_2 \in ST_2} Q_{\pi_2}.(r(\pi_2)-g) \geq \bar{0}.$$

However for π_2 optimal in Γ we have $Q_{\pi_2}.r(\pi_2) = g$ and $Q_{\pi_2}.g = g$, so $Q_{\pi_2}(r(\pi_2)-g) = \bar{0}$, proving the lemma. □

THEOREM 5.2. *For a two-person zero-sum finite undiscounted stochastic game, the following two assertions are equivalent:*

(i) *The game has value* g *and both players have optimal stationary strategies.*

(ii) (a) $g = \mathrm{Val}(\langle P(g)\rangle)$.

(b) *Let* $OE_i(k)$ *be the finite set of extreme optimal actions for player* i *in the matrix game* $P_k(g)$, $i \in \{1,2\}$. *Then there exist vectors* $v_1 \in \mathbb{R}^N$ *and* $v_2 \in \mathbb{R}^N$ *such that for all* $k \in S$

$$v_1(k) + g(k) = \operatorname*{Val}_{OE_1(k)\times A_2(k)} (\langle r_k + P_k(v_1)\rangle)$$

and

$$v_2(k) + g(k) = \operatorname*{Val}_{A_1(k)\times OE_2(k)} (\langle r_k + P_k(v_2)\rangle)$$

($\operatorname{Val}_{W\times Z}(\langle . \rangle)$ *means, that for the matrix game* $\langle . \rangle$ *the sets of pure action for the player* 1 *and* 2 *are respectively* W *and* Z.)

PROOF. Suppose (i) is true. Application of the lemma's 2.3 and 2.4 yields $g = g^*$ and $g = \mathrm{Val}(\langle P(g)\rangle)$. Let $OE_1(k)$ be the finite set of extreme optimal actions in the matrix game $\langle P_k(g)\rangle$ for player 1.

Let π_1^* be an optimal stationary strategy for player 1. By assumption we have $\min_{\pi_2} V(\pi_1^*, \pi_2) = g$, so by the lemma's 3.2 and 3.1 we can deduce

$$g = \min(\langle P_{\pi_1^*}(g)\rangle),$$

where $\langle P_{\pi_1^*}(g)\rangle$ equals an N-vector with k-th component equal to the function $P^k_{\pi_1^*}(g)\colon A_2^1(k) \to \mathbb{R}$, defined as

$$P^k_{\pi_1^*}(g)(a_2) = \sum_{\ell=1}^{N} P(\ell \mid k, \pi_1^*, a_2).g(\ell).$$

But this implies, that π_{1k}^* is optimal in the matrix game $\langle P_k(g)\rangle$, so

$$\pi_1^* \in \mathop{X}_{k=1}^{N} P(OE_1(k)). \tag{3}$$

Consider now the stochastic game, called Γ', in which the difference with the original stochastic game is that (a) for each state k player 1 has the

set $OE_1(k)$ as his pure action set and (b) the payoff in state k under pure actions $a_1 \in OE_1(k)$ and $a_2 \in A_2(k)$ equals $r(k,a_1,a_2)-g(k)$.

Lemma 5.1 shows that

$$\min_{\pi_2 \in ST_2} Q_{\pi_1^* \pi_2} (r(\pi_1^*,\pi_2)-g) = \bar{0}. \tag{4}$$

If π_2^* is optimal for player 2 in the original game, then

$$\max_{\pi_1 \in \underset{1}{\overset{N}{X}} P(OE_1(k))} Q_{\pi_1 \pi_2^*} r(\pi_1,\pi_2^*) = g,$$

so again using lemma 5.1 yields

$$\max_{\pi_1 \in \underset{1}{\overset{N}{X}} P(OE_1(k))} Q_{\pi_1 \pi_2^*} (r(\pi_1,\pi_2^*)-g) = \bar{0}. \tag{5}$$

Combining (4) and (5) gives that Γ' has value $\bar{0}$ and both players have optimal stationary strategies. But then we may apply theorem 4.4 yielding the existence of a vector v_1, such that for all $k \in S$:

$$v_1(k) = \operatorname{Val}_{OE_1(k) \times A_2(k)} (\langle r_k - g(k) + P_k(v_1) \rangle)$$

and this is equivalent to the first equation of (ii). The second equation can be derived analogously.

Suppose now (ii) is true. Let the stationary strategy $\pi_1^* \in \underset{k=1}{\overset{N}{X}} P(OE_1(k))$ be such that π_{1k}^* is an optimal action for player 1 in the matrix game $\langle r_k + P_k(v_1) \rangle$. Let π_2 be an arbitrary stationary strategy for player 2, then we have

$$v_1 + g \leq r(\pi_1^*,\pi_2) + P_{\pi_1^* \pi_2} . v_1. \tag{6}$$

Multiplying both sides of (6) by $Q_{\pi_1^* \pi_2}$ and rearranging terms yields

$$Q_{\pi_1^* \pi_2} g \leq Q_{\pi_1^* \pi_2} . r(\pi_1^*,\pi_2). \tag{7}$$

But as $\pi^*_{1k} \in P(OE_1(k))$, so π^*_{1k} is optimal in $\langle P_k(g)\rangle$ we have $P_{\pi^*_1\pi_2} g \geq g$ and by iterating we get $Q_{\pi^*_1\pi_2} g \geq g$. Inserting this last inequality in (7) yields the desired result

$$\min_{\pi_2 \in ST_2} Q_{\pi^*_1\pi_2}.r(\pi^*_1,\pi_2) \geq g. \tag{8}$$

In the same way we can show the existence of a stationary strategy π^*_2 for player 2, such that

$$\max_{\pi_1 \in ST_1} Q_{\pi_1\pi^*_2}.r(\pi_1,\pi^*_2) \leq g. \tag{9}$$

Remembering lemma 3.2 we see that the combination of (8) and (9) assures us the validity of (i). □

COROLLARY 5.3. *Let for a two-person zero-sum finite undiscounted stochastic game* (i) *or* (ii) *of theorem* 5.2 *hold. Let for each* $k \in S$ *the finite sets* $B_1(k)$ *and* $B_2(k)$ *be such that* $P(A_i(k)) \supset P(B_i(k)) \supset P(OE_i(k))$, $i \in \{1,2\}$, $k \in S$. *Then there exist a vector* $v \in \mathbb{R}^N$, *such that for each* $k \in S$

$$v(k) + g(k) = \operatorname{Val}_{B_1(k)\times B_2(k)} (\langle r_k + P_k(v)\rangle).$$

PROOF. Analogously as in the first part of the proof of theorem 5.2 it can be shown, that the undiscounted stochastic game, which differs from the original stochastic game by (a) the set of pure actions for player i in state k is $B_i(k)$, $i \in \{1,2\}$, $k \in S$ and (b) the payoff in state k for the actions $a_1 \in B_1(k)$ and $a_2 \in B_2(k)$ equals $r(k,a_1,a_2) - g(k)$, has value 0. Then again theorem 4.4 assures the corollary. □

Theorem 7.3.3 part (a) of FEDERGRUEN [5] is a special case (namely $B_i(k)=OE_i(k)$) of corollary 5.3.

We are now going to characterize the sets of optimal stationary strategist.

THEOREM 5.4. *Let for a two-person zero-sum finite undiscounted stochastic*

game (i) *or* (ii) *of theorem* 5.3 *hold. Let* $OE_1(k)$, $k \in S$ *be as in theorem* 5.2 *and let for a stationary strategy* π_1 *of player* 1 $R^*(\pi_1)$ *as in section* 3. *Then: a stationary strategy* π_1 *is optimal for player* 1 *if and only if*

(a) $\pi_1 \in \underset{k=1}{\overset{N}{X}} P(OE_1(k))$ *and*

(b) *there exist a vector* $v_1 \in \mathbb{R}^N$, *such that*

$$v_1(k) + g(k) \leq \underset{OE_1(k)\times A_2(k)}{\text{Val}} (<r_k+p_k(v_1)), \quad k \in S$$

and such that for each $k \in R^*(\pi_1)$ *the action* π_{1k} *assures exactly* $v_1(k)+g(k)$ *in the matrix game* $<r_k+p_k(v_1)>$.

PROOF. Let π_1 be an optimal stationary strategy for player 1. In the first part of the proof of theorem 5.2 we have proven part (a) of the theorem (see (*)). π_1 is optimal and also player 2 has optimal stationary strategies, so

$$\min_{\pi_2 \in ST_2} Q_{\pi_1\pi_2} r(\pi_1,\pi_2) = g.$$

By lemma 5.1 this gives

$$\min_{\pi_2 \in ST_2} Q_{\pi_1\pi_2} (r(\pi_1,\pi_2)-g) = \bar{0}.$$

As theorem 4.4 can also be applied to Markov decision problems, it now follows that there exist a vector $v_1 \in \mathbb{R}^N$ such that

$$v_1 = \min_{\pi_2 \in ST_2} (<r-g+P_{\pi_1\pi_2}.v_1>)$$

or equivalently

$$v_1 + g = \min_{\pi_2 \in ST_2} (<r+P_{\pi_1\pi_2}.v_1>). \tag{10}$$

As $\pi_1 \in \underset{k=1}{\overset{N}{X}} P(OE_1(k))$ (10) assures us that v_1 obeys

$$v_1(k) + g(k) \leq \underset{OE_1(k)\times A_2(k)}{\text{Val}} (<r_k+P_k(v_1)>) \tag{11}$$

(10) and (11) show the validity of (b).

Let π_1 and v_1 be such that (a) and (b) are true. Let $\tilde{\pi}_2$ be a stationary strategy for player 2, which is a best answer against π_1, clearly

$$Q_{\pi_1\tilde{\pi}_2}.r(\pi_1,\tilde{\pi}_2) \leq g \tag{12}$$

(see lemma 2.3).

By assumption $v_1(k)+g(k) \leq r(k,\pi_1,\tilde{\pi}_2) + P_{\pi_1\tilde{\pi}_2}.v_1$ for $k \in R_{\pi_1\tilde{\pi}_2}$. Multiplying this inequality by $Q_{\pi_1\tilde{\pi}_2}$ yields:

$$(Q_{\pi_1\tilde{\pi}_2}.r(\pi_1,\tilde{\pi}_2))(k) \geq (Q_{\pi_1\tilde{\pi}_2}.g)(k) = g(k), \quad k \in R_{\pi_1\tilde{\pi}_2} \tag{13}$$

Combining (12) and (13) gives

$$(Q_{\pi_1\tilde{\pi}_2}.r(\pi_1,\tilde{\pi}_2))(k) = g(k), \quad k \in R_{\pi_1\tilde{\pi}_2} \tag{14}$$

As $P_{\pi_1\tilde{\pi}_2}\, g \geq g$ (π_{1k} optimal in $\langle P_k(g)\rangle$, $\forall k \in S$) and as (14) holds for the recurrent states of π_1 and $\tilde{\pi}_2$ it follows from DENARDO and FOX [3] (lemma 4), that also $(Q_{\pi_1\tilde{\pi}_2}.r(\pi_1,\tilde{\pi}_2))(k) \geq g(k)$ if k is a transient state and the combination with (12) gives

$$(Q_{\pi_1\tilde{\pi}_2}.r(\pi_1,\tilde{\pi}_2))(k) = g(k) \tag{15}$$

for k transient. (14) and (15) together with the assumption on $\tilde{\pi}_2$ show the optimality of π_1.

Of course a similar theorem can be stated for player 2.

The next theorem characterizes the sets of optimal stationary strategies in games, where for each pair of stationary strategies all states are recurrent. HOFFMAN and KARP [6] were the first, who showed, that these games do have a value and that both players have optimal stationary strategies.

THEOREM 5.5. *Let for a two-person zero-sum finite undiscounted stochastic*

game under each pair of stationary strategies all states be recurrent. Then a stationary strategy π_1 *for player* 1 *is optimal if and only if*

(a) $\pi_1 \in \underset{k=1}{\overset{N}{X}} P(OE_1(k))$ ($OE_1(k)$ *be as in theorem* 5.2)

(b) *there exist a vector* $v_1 \in \mathbb{R}^N$, *such that*

$$v_1(k) + g(k) = \underset{OE_1(k)\times A_2(k)}{\text{Val}} (<r_k + P_k(v_1)>), \quad k \in S$$

and such that π_{1k} *is optimal in the matrix game* $(<r_k + P_k(v_1)>)$ *for* $\forall k \in S$.

PROOF. Looking at theorem 5.4 it can be seen that the only thing we should prove is that in the inequality in (b) the inequality sign cannot be strict in any component.

So suppose it is strict in one or more components, then there exists a $\tilde{\pi}_1 \in \underset{k=1}{\overset{N}{X}} P(OE_1(k))$ such that for every $\pi_2 \in ST_2$ it holds.

$$v_1 + g \leq r(\tilde{\pi}_1, \pi_2) + P_{\tilde{\pi}_1 \pi_2} . v_1$$

with strict inequality in say state k. Multiplying by $Q_{\tilde{\pi}_1 \pi_2}$ yields (all states recurrent):

$$Q_{\tilde{\pi}_1 \pi_2} . r(\tilde{\pi}_1, \pi_2) \geq g$$

with strict inequality in at least one component. As then, this also holds for the best reply of player 2 against $\tilde{\pi}_1$, it follows that $\tilde{\pi}_1$ is a stationary strategy for player 1 which assures him in at least 1 start more than g, contradicting lemma 2.4. □

REFERENCES

1. BEWLEY, T. & E. KOHLBERG, *The asymptotic theory of stochastic games*, Math. of O.R., vol 1, pp 197-308, 1976.

2. BEWLEY, T. & E. KOHLBERG, *On stochastic games with stationary optimal strategies*, Math. of O.R., vol 3, pp 104-125, 1978.

3. DENARDO, E. & B. FOX, *Multichain Markov Renewal Program*, SIAM J. Appl. Math. 16, pp 477-496, 1968.

4. DERMAN, C., *Finite State Markovian Decision Processes*, Academic Press, New York.

5. FEDERGRUEN, A., *Markovian Control Problems*, Ph.D. dissertation, Mathematisch Centrum, Amsterdam, 1978.

6. HOFFMAN, A. & R. KARP, *On non-terminating stochastic games*, Man. Sc. 12, pp 359-370, 1966.

7. SHAPLEY, L., *Stochastic games*, Proc. Nat. Acad. Sci. U.S.A. 39, pp 1095-1100, 1953.