stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 110/79 JULI
(DEPARTMENT OF OPERATIONS RESEARCH)

S.H. TIJS & O.J. VRIEZE

CHARACTERIZING PROPERTIES OF THE VALUE FUNCTION OF STOCHASTIC GAMES

Preprint

2e boerhaavestraat 49 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).

1980 Mathematics subject classification: 90D05,90D15

Characterizing properties of the value function of stochastic games[*)]

by

S.H. Tijs[**)] & O.J. Vrieze

ABSTRACT

The axiomatic characterization of the value function of two-person zero-sum games in normal form by VILKAS and TIJS is extended to dynamic games. Special attention is given to the value function of discounted two-person zero-sum stochastic games. Furthermore, value functions for stochastic games with arbitrary evaluation rules and general state and action spaces are examined.

KEY WORDS & PHRASES: *Stochastic games, value of stochastic games, characterization of the value function.*

---

*) This report will be submitted for publication elsewhere.

**) Department of Mathematics, Catholic University, Nijmegen, The Netherlands.

1. INTRODUCTION

E.I. VILKAS characterized in [4] the value function for matrix games in an axiomatic way. TIJS extended in [3] this idea to the characterization of the value function for finite linear programs and to the characterization of the value function on the class of all determined two-person zero-sum games in normal form.

In this paper we give an axiomatic description of the value function of discounted two-person zero-sum stochastic games. This will be done in Section 3, where the characterizing axioms can be indicated by the terms: objectivity, monotony and sufficiency for both players (or sufficiency for one of the players and symmetry). Herewith we will use the concept of a superfluous action for a player in a state of the game. Such an action can be ignored by a player without his being punished, because the value of the game remains invariant. We will also show in Section 3 that the characterizing axioms are mutually independent.

In Section 4 we introduce the concept of a weak superfluous action. Ignoring such an action again does not change the value. With this concept we are able to characterize the value function without using the monotony axiom. Next in Section 4 a general description of the value functions of more general stochastic games with arbitrary monotone evaluation rules is given.

In Section 2 we expose the model, which will be examined, we introduce the necessary concepts and give two lemmas, which contain some well-known facts about discounted two-person zero-sum stochastic games.

2. PRELIMINARIES

For a finite set $B = \{1,2,\dots,h\}$ we denote by $\mathcal{P}(B)$ the set of probability measures on B. The elements of $\mathcal{P}(B)$ will be identified in an obvious way with elements of the set

$$\{x = (x_1,x_2,\dots,x_h) \in \mathbb{R}^h;\ x_k \geq 0 \text{ for all } k \in B \text{ and } \sum_{k=1}^{h} x_k = 1\}.$$

A discounted two-person zero-sum stochastic game can be characterized by a six-tuple

$$\Gamma = \langle S, \{A_1(k): k \in S\}, \{A_2(k): k \in S\}, r, P, \beta \rangle,$$

where

- $S = \{1,2,\ldots,N\}$,
- $A_1(k) = \{1,2,\ldots,m_k\}$, $A_2(k) = \{1,2,\ldots,n_k\}$ for $k \in S$,
- r is a real-valued function defined on the set of triples
  $T = \{(k,i,j): k \in S,\ i \in A_1(k),\ j \in A_2(k)\}$,
- P is a map from T into the set $\mathcal{P}(S)$,
- $\beta \in (0,1)$.

S will be called the state space, $A_i(k)$ the set of pure actions of player i ($i = 1,2$) in state k, r the reward function for player 1 ($-r$ the reward function for player 2), P the transition probability map and β the discount factor.

Such a stochastic game corresponds with a dynamic system, where the dynamic behavior 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 stage t the system is in state k and if player 1 selects action $i \in A_1(k)$ and player 2 action $j \in A_2(k)$, then two things happen:

(1) player 1 obtains an immediate reward $r(k,i,j)$ from player 2;

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

A history dependent strategy π_i for player i in the game Γ ($i = 1,2$) is a rule, which, for each stage $t \in \{0,1,2,\ldots\}$ and each state $k \in S$, selects, dependent on t and the history of the game at stage t, an element of $\mathcal{P}(A_i(k))$ (according to which probability measure player i has to choose his pure action in that situation, if he adopts that strategy). The history of the game at stage t is the sequence of states and actions, that actually have occurred up to time t. The set of history dependent strategies for player i is denoted by H_i.

A stationary strategy for player i is a rule, where for each stage t and each state $k \in S$ the selection of an element of $\mathcal{P}(A_i(k))$ is made independent of t and the history of the game at time t. A stationary strategy can be denoted by $\pi_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{iN})$, where $\pi_{ik} \in \mathcal{P}(A_i(k))$ for all $k \in S$, i.e., each time the system is in state k, player i will play the mixed action π_{ik}. The set of stationary strategies for player i will be denoted by ST_i.

If the players 1 and 2 play strategies $\pi_1 \in H_1$ and $\pi_2 \in H_2$, respectively, then for each time t the expected reward at time t can be computed, and will be denoted by the N-vector

$$r_t(\pi_1, \pi_2) = (r_t(\pi_1, \pi_2)(1), \ldots, r_t(\pi_1, \pi_2)(N)),$$

where the k-th coordinate corresponds with the specific game with state k as starting state. In stochastic games with discount factor β this stream of expected payoffs is evaluated by discounting the expected payoff at time t by the factor β^t. The total discounted expected payoff, denoted by

$$v(\pi_1, \pi_2) = (v(\pi_1, \pi_2)(1), \ldots, v(\pi_1, \pi_2)(N))$$

equals then

$$v(\pi_1, \pi_2) = \sum_{t=0}^{\infty} \beta^t r_t(\pi_1, \pi_2),$$

where again the k-th component of $v(\pi_1, \pi_2)$ corresponds to the specific game with state k as the starting state. Since

$$\| r_t(\pi_1, \pi_2) \|_\infty \leq M$$

with

$$M = \max_{(k,i,j) \in T} |r(k,i,j)|,$$

it is obvious that $v(\pi_1, \pi_2)$ is well-defined for each pair $(\pi_1, \pi_2) \in H_1 \times H_2$. Of course, we suppose that player 1 wants to maximize the coordinates of $v(\pi_1, \pi_2)$ and player 2 wants to minimize this vector.

DEFINITION 2.1. A discounted stochastic game is said to have a *value* if

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

DEFINITION 2.2. If a discounted stochastic game has a value, say V, then, for each $\varepsilon \geq 0$, a strategy for player 1 is called ε-*optimal*, if

$$\inf_{\pi_2 \in H_2} v(\pi_1,\pi_2) \geq V - \varepsilon \underline{1}.$$

(Here $\underline{1}$ is the vector in $\mathbb{R}^N$, for which all coordinates are equal to 1.) A strategy $\pi_2 \in H_2$ is called ε-*optimal* for player 2, if

$$\sup_{\pi_1 \in H_1} v(\pi_1,\pi_2) \leq V + \varepsilon \underline{1}.$$

0-optimal strategies are usually called *optimal* strategies.

SHAPLEY was the first who analyzed discounted stochastic games and proved the following lemma [2].

LEMMA 2.3. *Let Γ be a β-discounted stochastic game (with finite state space and finite pure action sets). Then Γ possesses a value, say $V(\Gamma)$. Furthermore, both players possess optimal stationary strategies. For $i = 1,2$, a stationary strategy $\pi_i = (\pi_{i1},\ldots,\pi_{iN})$ is optimal for player i, if and only if for each $k \in S$, the action π_{ik} is an optimal action for player i in the matrix game*

$$\langle r(k,\cdot,\cdot) + \beta \sum_{\ell \in S} p(\ell|k,\cdot,\cdot)V(\Gamma)(\ell)\rangle,$$

i.e., the $m_k \times n_k$-matrix game with in the (i,j)-th cell of the matrix the real number $r(k,i,j) + \beta \sum_{\ell \in S} p(\ell|k,i,j)V(\Gamma)(\ell)$. This matrix game has value $V(\Gamma)(k)$.

The following lemma, well-known in Markov decision theory, will be useful in the sequel. A proof can be found in DENARDO & FOX [1], p.474.

LEMMA 2.4. *For $(\pi_1,\pi_2) \in ST_1 \times ST_2$ the total β-discounted expected payoff $v(\pi_1,\pi_2)$ is the unique solution of the following set of N functional equations in $x \in \mathbb{R}^N$:*

$$x_k = r(k,\pi_{1k},\pi_{2k}) + \beta \sum_{\ell\in S} p(\ell|k,\pi_{1k},\pi_{2k})x_\ell \quad \text{for all } k \in S,$$

where $r(k,\pi_{1k},\pi_{2k})$ *and* $p(\ell|k,\pi_{1k},\pi_{2k})$ *are the expectations of the functions* $r(k,\cdot,\cdot)$ *and* $p(\ell|k,\cdot,\cdot)$ *on* $A_1(k)\times A_2(k)$ *with respect to the probability measure* $\pi_{1k}\times\pi_{2k}$. *If* $y \in \mathbb{R}^N$ *is such that for all* $k \in S$

$$y_k \le (\ge)\ r(k,\pi_{1k},\pi_{2k}) + \beta \sum_{\ell\in S} p(\ell|k,\pi_{1k},\pi_{2k})y_\ell,$$

then

$$y \le v(\pi_1,\pi_2) \qquad (y \ge v(\pi_1,\pi_2)).$$

DEFINITION 2.5. For a β-discounted stochastic game Γ an action $i \in A_1(k)$ $(k \in S)$, is called *superfluous*, if there exists an action $\bar\pi_{1k} \in \mathcal{P}(A_1(k))$ such that $\bar\pi_{1k}(i) = 0$ and such that for each $j \in A_2(k)$:

$$r(k,i,j) + \beta \sum_{\ell\in S} p(\ell|k,i,j)V(\Gamma)(\ell) \le \tag{2.1}$$
$$r(k,\bar\pi_{1k},j) + \beta \sum_{\ell\in S} p(\ell|k,\bar\pi_{1k},j)V(\Gamma)(\ell).$$

If (2.1) holds and $\bar\pi_{1k}(i) = 0$, then we say that i is superfluous in view of action $\bar\pi_{1k}$. An action $j \in A_2(k)$ $(k \in S)$ is called superfluous if there exists an action $\bar\pi_{2k} \in \mathcal{P}(A_2(k))$ such that $\bar\pi_{2k}(j) = 0$ and such that for each $i \in A_1(k)$:

$$r(k,i,j) + \beta \sum_{\ell\in S} p(\ell|k,i,j)V(\Gamma)(\ell) \ge$$
$$r(k,i,\bar\pi_{2k}) + \beta \sum_{\ell\in S} p(\ell|k,i,\bar\pi_{2k})V(\Gamma)(\ell).$$

In the next section it is shown, that ignoring superfluous actions, does not disturb the value of the game.

In the following we denote by $G(N,\beta)$ the family of β-discounted stochastic games with N states and finite sets of actions for both players. We want to characterize the function $V: G(N,\beta) \to \mathbb{R}^N$, where for each $\Gamma \in G(N,\beta)$, $V(\Gamma)$ is the value of the game Γ.

3. CHARACTERIZING PROPERTIES OF THE VALUE FUNCTION FOR DISCOUNTED STOCHASTIC GAMES

In this section we first prove four properties of the value function on $G(N,\beta)$. Next we show that a function on $G(N,\beta)$, satisfying these four properties, necessarily must be the value function. Furthermore, we show that the four properties are mutually independent.

PROPERTY 1 (Objectivity). Let $\Gamma \in G(N,\beta)$ be such that for state $k \in S$ we have $m_k = n_k = 1$ and $p(k|k,1,1) = 1$. Then $V(\Gamma)(k) = (1-\beta)^{-1}r(k,1,1)$.

PROOF. If we start in state k, then we stay in state k for all pairs of strategies (π_1,π_2) and the total β-discounted payoff equals $\sum_{t=0}^{\infty} \beta^t r(k,1,1) = (1-\beta)^{-1}r(k,1,1)$. □

PROPERTY 2 (Monotony). Let $\Gamma',\Gamma'' \in G(N,\beta)$ and suppose that all game parameters, except the reward functions r' and r", are the same. Let $r'(k,i,j) \le r''(k,i,j)$ for all $(k,i,j) \in \Gamma$. Then $V(\Gamma') \le V(\Gamma'')$.

PROOF. Trivial. □

PROPERTY 3.1 (Sufficiency for player 1). For $\Gamma \in G(N,\beta)$, let an action $i \in A_1(k)$ be superfluous. Let $\Gamma' \in G(N,\beta)$ be the game, which results from Γ, when we delete action i. Then $V(\Gamma) = V(\Gamma')$.

PROOF. First we note that obviously $V(\Gamma') \le V(\Gamma)$ (the set of strategies for player 1 is reduced, the set of strategies for player 2 and the other game parameters remain unchanged; now the inequality can directly be seen from the definition of a value). Hence we have to prove that also $V(\Gamma') \ge V(\Gamma)$.

Suppose that we can construct a stationary strategy $\tilde{\pi}_1$ for player 1 in the game Γ', satisfying

$$(3.1) \qquad r(s,\tilde{\pi}_{1s},\pi_{2s}) + \beta \sum_{\ell\in S} p(\ell|s,\tilde{\pi}_{1s},\pi_{2s})V(\Gamma)(\ell) \ge V(\Gamma)(s)$$

for all $s \in S$ and all stationary strategies π_2 in the game Γ'. (Note that ST_2 is the same set in both games Γ and Γ'.)

Then, in view of Lemma 2.4 with $V(\Gamma)$ in the role of y, we may conclude that

$V(\Gamma) \le V(\tilde{\pi}_1,\pi_2)$ for all $\pi_2 \in ST_2$. But as Lemma 2.3 shows that player 2 has an optimal stationary strategy, we have

$$V(\Gamma) \le \inf_{\pi_2 \in ST} V(\tilde{\pi}_1,\pi_2) \le V(\Gamma').$$

Hence, the only thing required to finish the proof is to construct $\tilde{\pi}_1 = (\tilde{\pi}_{11},\dots,\tilde{\pi}_{1N})$ with property (3.1). For each $s \in S - \{k\}$ take $\tilde{\pi}_{1s}$ such that $\tilde{\pi}_{1s}$ is optimal in the matrix game $<r(s,\cdot,\cdot) + \beta \sum_{\ell \in S} p(\ell|s,\cdot,\cdot)V(\Gamma)(\ell)>$. Since this matrix game has value $V(s)$ (Lemma 2.3), property (3.1) is satisfied for $s \neq k$. Now let $i \in A_1(k)$ be superfluous in view of action $\bar{\pi}_{1k} \in \mathcal{P}(A_1(k))$, i.e., (2.1) holds and $\bar{\pi}_{1k}(i) = 0$. Take $\pi^*_{1k} \in \mathcal{P}(A_1(k))$ such that π^*_{1k} is optimal in the matrix game $<r(k,\cdot,\cdot) + \beta \sum_{\ell \in S} p(\ell|k,\cdot,\cdot)V(\Gamma)(\ell)>$. Now define $\tilde{\pi}_{1k}$ as follows:

$$\tilde{\pi}_{1k}(i') = \pi^*_{1k}(i') + \pi^*_{1k}(i)\bar{\pi}_{1k}(i')$$

for all i' in the set of actions $A_1(k) - \{i\}$ of player 1 in state k of the game Γ'.

Then, obviously, $\tilde{\pi}_{1k} \in \mathcal{P}(A_1(k) - \{i\})$. Furthermore, it follows with the aid of (2.1) that (3.1) is also satisfied for $s = k$. This completes the proof of property 3.1. □

Quite analogously, we can show the validity of:

PROPERTY 3.2 (Sufficiency for player 2). For $\Gamma \in G(N,\beta)$, let an action $j \in A_2(k)$ be superfluous. Let $\Gamma' \in G(N,\beta)$ be the game, which results from Γ, when we delete action j. Then $V(\Gamma') = V(\Gamma)$.

Now we state our main result.

THEOREM 3.1. *A function* $f\colon G(N,\beta) \to \mathbb{R}^N$ *equals the value function if and only if* f *obeys the following axioms:*

Axiom 1 (objectivity). *If* $\Gamma \in G(N,\beta)$ *is such that for state* $k \in S$ *we have* $m_k = n_k = 1$ *and* $p(k|k,1,1) = 1$, *then* $f(\Gamma)(k) = (1-\beta)^{-1}r(k,1,1)$.

Axiom 2 (monotony). *If two games* Γ' *and* Γ'' *in* $G(N,\beta)$ *of equal size are such that the transition probability maps are the same and for the reward* *functions* r' *and* r" *we have* $r' \le r''$, *then* $f(\Gamma') \le f(\Gamma'')$.

Axiom 3.i (Sufficiency for player i, i = 1,2). *If* $\Gamma' \in G(N,\beta)$ *results from* $\Gamma \in G(N,\beta)$ *by deleting a superfluous action of player* i, *then* $f(\Gamma') = f(\Gamma)$.

PROOF. The "only if" part of the theorem follows from the already proven properties 1, 2, 3.1 and 3.2 for the value function.
Now let f be a function, obeying the axioms 1, 2, 3.1 and 3.2.
The proof of the "if" part of the theorem will proceed in two steps. In the first step we only consider stochastic games $\Gamma \in G(N,\beta)$, which have a state k and a pair of actions $(i_0,j_0) \in A_1(k) \times A_2(k)$ such that

$$p(k|k,i,j) = 1 \quad \text{if} \quad i = i_0 \quad \text{or} \quad j = j_0$$

and such that

$$(3.2) \qquad \inf_{j \in A_2(k)} r(k,i_0,j) = r(k,i_0,j_0) = \sup_{i \in A_1(k)} r(k,i,j_0).$$

For such games we show that

$$(3.3) \qquad f(\Gamma)(k) = V(\Gamma)(k).$$

In the second step we consider arbitrary elements of $G(N,\beta)$ and connect them with games of the form treated in the first step.

(1) Let Γ be a game in $G(N,\beta)$ with state k and actions i_0 and j_0 as indicated above. Obviously $V(\Gamma)(k) = (1-\beta)^{-1} r(k,i_0,j_0)$.
Let M be a large real number. Look at the games Γ' and Γ'' in $G(N,\beta)$ which differ from Γ only in the reward functions r' and r" as follows:

$$r'(\ell,i,j) = r(\ell,i,j) \quad \text{if} \quad \ell = k \quad \text{and} \quad i = i_0 \quad \text{or} \quad j = j_0$$
$$r'(\ell,i,j) = r(\ell,i,j) - M \text{ elsewhere.}$$
$$r''(\ell,i,j) = r(\ell,i,j) \quad \text{if} \quad \ell = k \quad \text{and} \quad i = i_0 \quad \text{or} \quad j = j_0,$$
$$r''(\ell,i,j) = r(\ell,i,j) + M \text{ elsewhere.}$$

From the monotony properties of f follows

$$(3.4) \qquad f(\Gamma') \le f(\Gamma) \le f(\Gamma'').$$

We concentrate our attention on the game Γ'. For M large enough it is obvious that for each $i \in A_1(k) - \{i_0\}$ we have

$$r'(k,i,j) + \beta \sum_{\ell \in S} p'(\ell|k,i,j)V(\Gamma')(\ell) \le$$

$$r'(k,i_0,j) + \beta\, p'(k,i_0,j)V(\Gamma')(k) \quad \text{for all } j \in A_2(k).$$

(Note that $V(\Gamma)(k) = V(\Gamma')(k)$.) This shows that for the game Γ' each action i in state k, unequal to i_0, is superfluous in view of action i_0. So we may successively delete all actions $i \neq i_0$, without disturbing the f-value by axiom 3.1. Then there only remains action i_0 in state k. In view of (3.2) we can in this new game also delete each action $j \in A_2(\ell) - \{j_0\}$, using axiom 3.2. This results in a game $\tilde{\Gamma}$ with $f(\Gamma') = f(\tilde{\Gamma})$, and where state k is such that $\tilde{m}_k = \tilde{n}_k = 1$ and $\tilde{p}(k|k,1,1) = 1$. In view of axiom 1 we obtain

$$f(\Gamma')(k) = f(\tilde{\Gamma})(k) = (1-\beta)^{-1}\tilde{r}(k,1,1) =$$

$$(1-\beta)^{-1}r'(k,i_0,j_0) = (1-\beta)^{-1}r(k,i_o,j_0) = V(\Gamma)(k).$$

Analogously it can be shown that also $f(\Gamma'')(k) = (1-\beta)^{-1}r(k,i_0,j_0) = V(\Gamma)(k)$. Combining these results with (3.4) leads to (3.3).

(2) Now we start with the second step of the proof. Let $\Gamma \in G(N,\beta)$ be an arbitrary game with value $V(\Gamma)$ and let $k \in S$ be an arbitrary state. Now consider the game Γ^k, which is constructed from Γ, by adding in state k an action i_0 for player 1 and an action j_0 for player 2 and by extending the reward function r and the transition probability map P of Γ such that

$r(k,i,j) = (1-\beta)V(\Gamma)(k)$ if $i = i_0$ or $j = j_0$, and

$p(k|k,i,j) = 1$ if $i = i_0$ or $j = j_0$.

Clearly, $V(\Gamma^k) = V(\Gamma)$. Furthermore, Γ^k is a game with state k of the type treated in the first step of the proof. Hence,

$$f(\Gamma^k)(k) = (1-\beta)^{-1}r(k,i_0,j_0) = V(\Gamma)(k). \tag{3.5}$$

Our proof is complete, if we can show that the actions i_0 and j_0 are superfluous, because then axioms 3.1 and 3.2 give

$$f(\Gamma^k)(k) = f(\Gamma)(k).$$

Take an optimal stationary strategy π_1 for player 1 in the game Γ. By Lemma 2.3 we have

$$(3.6) \qquad r(k,\pi_{1k},j) + \beta \sum_{\ell \in S} p(\ell|k,\pi_{1k},j)V(\Gamma)(\ell) \geq V(\Gamma)(k)$$

$$\text{for each } j \in A_2(k).$$

The same holds for the game Γ^k. For the action j_0 we have

$$(3.7) \qquad r(k,\pi_{1k},j_0) + \beta \sum_{\ell \in S} p(\ell|k,\pi_{1k},j)V(\Gamma)(\ell) =$$

$$(1-\beta)V(\Gamma)(k) + \beta\, V(\Gamma)(k) = V(\Gamma)(k).$$

However, for the right-hand sides in (3.6) and (3.7) the following holds:

$$(3.8) \qquad V(\Gamma)(k) = r(k,i_0,j) + \beta\, p(k|k,i_0,j)V(\Gamma)(k) =$$

$$= r(k,i_0,j) + \beta \sum_{\ell \in S} p(\ell|k,i_0,j)V(\Gamma)(\ell)$$

$$\text{for all } j \in A_2(k) \cup \{j_0\}.$$

Since $V(\Gamma^k) = V(\Gamma)$, it follows from (3.6), (3.7) and (3.8) that in the game Γ^k the action i_0 is superfluous for player 1 in view of π_{1k}. In an analogous way, one can show that j_0 is superfluous. This completes the proof of the theorem. □

Now we want to show that the four axioms 1, 2, 3.1 and 3.2 are independent. We do this by giving for each triple of them a function from $G(N,\beta)$ into $\mathbb{R}^N$ satisfying these three axioms and not the fourth axiom.

(a) (Objectivity). Let $f_1\colon G(N,\beta) \to \mathbb{R}^N$ be the map with $f_1(\Gamma) = 0$ for all $\Gamma \in G(N,\beta)$. Obviously, f_1 satisfies the axioms 2, 3.1 and 3.2, but not axiom 1.

(b) (Sufficiency for player 1). Let $f_2\colon G(N,\beta) \to \mathbb{R}^N$ be the map defined by

$$f_2(\Gamma)(k) = \min_{i,j}\{r(k,i,j) + \beta \sum_{\ell \in S} p(\ell|k,i,j)V(\Gamma)(\ell)\}$$

$$\text{for each } \Gamma \in G(N,\beta) \text{ and } k \in S.$$

Then f_2 obeys the axioms 1, 2 and 3.2, but not axiom 3.1.

(c) (Sufficiency for player 2). Let f_3: $G(N,\beta) \to \mathbb{R}^N$ be the map with

$$f_3(\Gamma)(k) = \max_{i,j}\{r(k,i,j) + \beta \sum_{\ell\in S} p(\ell|k,i,j)V(\Gamma)(\ell)\}.$$

Then f_3 satisfies the axioms 1, 2 and 3.1, but not axiom 3.2.

(d) (Monotony). More work has to be done to show that axiom 2 is independent of the other axioms. First we look at matrix games. For a matrix game A, let I(A) be the set of pure strategies (rows) i for player 1, which are not superfluous and which are such that $\pi_1(i) = 0$ for each optimal action π_1 of player 1. Let J(A) have the analoguous meaning for player 2. Suppose now that a superfluous row of A is deleted, resulting in a matrix game A'. Then it is obvious that $I(A') = I(A)$ and $J(A') \subset J(A)$. It may happen that $J(A') \neq J(A)$ as the following Example 3.2 shows. Now for a matrix game A let $I_0(A)$ be the subset of I(A), consisting of those i for which $i \in I(A')$ for each matrix A', which can be obtained from A by deleting superfluous rows and superfluous columns in any possible order. Let $J_0(A)$ have the analoguous meaning for player 2. In Example 3.2 the sets $I_0(A)$ and $J_0(A)$ are empty. That is not the case in Example 3.3.

<u>EXAMPLE 3.2</u>. Let A be the matrix $\begin{bmatrix} 5 & 0 & 1 \\ 0 & 6 & 2 \\ 5 & 4 & 3 \end{bmatrix}$. Then $I(A) = \{2\}$ and $J(A) = \{1,2\}$. If the superfluous row 1 of A is deleted, we obtain the matrix $A' = \begin{bmatrix} 0 & 6 & 2 \\ 5 & 4 & 3 \end{bmatrix}$ for which $I(A') = \{1'\} = \{2\} = I(A)$ and $J(A') = \{1'\} = \{1\} \subset J(A)$. Furthermore, $I_0(A) = \emptyset$. $J_0(A) = \emptyset$.

<u>EXAMPLE 3.3</u>. Let B be the matrix $\begin{bmatrix} 6 & 0 & 1 \\ 0 & 6 & 2 \\ 5 & 4 & 3 \end{bmatrix}$. Then $I_0(B) = J_0(B) = \{1,2\}$.

Now we return to stochastic games. For $\Gamma \in G(N,\beta)$ and $k \in S$ let B_k be the matrix game $\langle r(k,\cdot,\cdot) + \beta \sum_{\ell\in S} p(\ell|k,\cdot,\cdot)V(\Gamma)(\ell)\rangle$. Let f_4: $G(N,\beta) \to \mathbb{R}^N$ be defined as follows:

$$f_4(\Gamma)(k) = V(\Gamma)(k) - \sum_{i\in I_0(B_k)} \sum_{j\in J_0(B_k)} b_k(i,j).$$

Then obviously f_4 satisfies the axioms 1, 3.1 and 3.2. We show that f_4 does not satisfy axiom 2.

Suppose that Γ is such that B_1 is equal to the matrix B in Example 3.3. (Such a game exists!) Then $V(\Gamma)(1) = 3$. Hence $f_4(\Gamma)(1) = 3-6-0-0-6 = -9$. Let Γ' be the stochastic game, which differs from Γ only in the fact that $r'(k,1,1) = r(k,1,1)+1$. Then $r \le r'$, and $V(\Gamma) = V(\Gamma')$, but $f_4(\Gamma')(1) = 3-7-0-0-6 = -10 < f_4(\Gamma)(1)$. Hence f_4 does not satisfy the monotony axiom.

Finally, we want to look at another interesting property of the value function, called *symmetry*. Therefore, we introduce the transpose of a stochastic game, which is the stochastic game that we obtain by interchanging the names of the players.

DEFINITION 3.4. Let $\Gamma = \langle S,\{A_1(k): k \in S\}, \{A_2(k): k \in S\}, r, P, \beta\rangle$ be a stochastic game. Then the transpose Γ^T of Γ is the stochastic game

$$\langle S,\{A_2(k): k \in S\},\{A_1(k): k \in S\},r^T,P^T,\beta\rangle,$$

where

$$r^T(k,a_2,a_1) = -r(k,a_1,a_2)$$

and

$$p^T(\ell|k,a_2,a_1) = p(\ell|k,a_1,a_2) \quad \text{for all } k,\ell \in S \text{ and } a_1 \in A_1(k),\ a_2 \in A_2(k).$$

Now we say that a function $f\colon G(N,\beta) \to \mathbb{R}^N$ is symmetric, if the following axiom holds:

Axiom 4 (symmetry). $f(-\Gamma^T) = -f(\Gamma)$ for all $\Gamma \in G(N,\beta)$.

It is straightforward to prove that the value function V has the following

PROPERTY 4 (Symmetry). $V(-\Gamma^T) = -V(\Gamma)$ for all $\Gamma \in G(N,\beta)$.

Furthermore, it is simple to show that axiom 3.2 follows from axiom 3.1 and axiom 4. This implies that we have the following alternative characterization of the value function.

THEOREM 3.5. *A function* $f\colon G(N,\beta) \to \mathbb{R}^N$ *equals the value function iff* f *obeys axioms* 1, 2, 3.1 *and* 4.

4. REMARKS AND GENERALIZATIONS

We start with giving another characterization of the value function, in which the monotony property no longer *plays* a role. For that purpose we need the following

DEFINITION 4.1. For a β-discounted stochastic game Γ an action i in state k for player 1 is called *weakly superfluous,* if for each $\pi_{1k} \in \mathcal{P}(A_1(k))$ there exists an action $\tilde{\pi}_{1k} \in \mathcal{P}(A_1(k))$ with $\tilde{\pi}_{1k}(i) = 0$ and such that

$$\inf_{j \in A_2(k)} r(k,\tilde{\pi}_{1k},j) + \beta \sum_{\ell \in S} p(\ell|k,\tilde{\pi}_{1k},j)V(\Gamma)(\ell) \geq$$

$$\inf_{j \in A_2(k)} r(k,\pi_{1k},j) + \beta \sum_{\ell \in S} p(\ell|k,\pi_{1k},j)V(\Gamma)(\ell).$$

It is obvious that a superfluous action is also a weakly superfluous action (but the converse does not necessarily hold). This implies that the following property is stronger than property 3.1.

PROPERTY 3.1w (Weak sufficiency for player 1). For a game Γ let action $i \in A_1(k)$ $(k \in S)$ be weakly superfluous. Let Γ' be the game, which results when action i is deleted. Then $V(\Gamma) = V(\Gamma')$.

It will be ovious how to formulate

PROPERTY 3.2w (Weak sufficiency for player 2).

Now we are ready to give the other characterization.

THEOREM 4.2. *A function* $f\colon G(N,\beta) \to \mathbb{R}^N$ *equals the value function if and only if* f *satisfies the objectivity property and the properties* 3.1w *and* 3.2w.

PROOF. In the proof of Theorem 3.1, the only place where the monotony axiom is used, is in the first step, where the game Γ is compared with two games Γ' and Γ''. From those last two games superfluous actions could be deleted. But now we no longer need games Γ' and Γ'', because directly in game Γ with a state k with a saddle-point as in the proof of Theorem 3.1, at once all actions of player 1 unequal to i_0 can be deleted successively, as they are all weakly superfluous in view of action i_0. The remainder of the proof proceeds analogously as the proof of Theorem 3.1. □

Now we also want to look at value functions for more general classes of dynamic games, where the evaluation of the payoff stream is not necessarily the β-discount criterion but, e.g., the t-step criterion, the total expected payoff criterion, or the average expected payoff criterion, and where the state and action spaces are not necessarily finite.

We look at dynamic games of the form

$$\Gamma = \langle S, \{A_1(s): \; s \in S\}, \{A_2(s): s \in S\}, r, P \rangle,$$

where the parameters have an analogous meaning as in Section 2, but now the state space and action spaces are non-empty measurable spaces, with measurable one-point sets, and the reward function and the transition probability map are measurable functions. Let us denote for a fixed state space S, the family of those games by G_S. It will be obvious what history dependent strategies are in such a game Γ. Now suppose that there is given some evaluation rule $w: H_1 \times H_2 \to \mathbb{R}^S$, which assigns to each pair $(\pi_1,\pi_2) \in H_1 \times H_2$ an element $w(\pi_1,\pi_2) \in \mathbb{R}^S$, where $w(\pi_1,\pi_2)(s)$ can be interpreted as the payoff of the specific game with state s as starting state, when the players choose the strategies π_1 and π_2. For these general stochastic games with prescribed evaluation rule w the notions of value and ε-optimal strategies can be defined similarly as in Section 2.

For two games Γ' and Γ" of equal size, which differ only in the reward functions r' and r", we write $\Gamma' \le \Gamma'' + d$ $(d \in \mathbb{R})$ if $r'(s,a_1,a_2) \le r''(s,a_1,a_2) + d$ for all $s \in S$ and $(a_1,a_2) \in A_1(s) \times A_2(s)$. In the following we only look at evaluation rules w which satisfy the following monotony condition:

Assumption M. There exists a $c \in (0,\infty]$ such that $\Gamma' \le \Gamma'' + d$ implies that

$$w'(\pi_1,\pi_2)(s) \le w''(\pi_1,\pi_2)(s) + dc \quad \text{for all } (\pi_1,\pi_2) \in H_1 \times H_2 \text{ and } s \in S.$$

Assumption M assures that $\Gamma' = \Gamma'' + d$ implies that

$$w'(\pi_1,\pi_2)(s) = w''(\pi_1,\pi_2)(s) + cd \quad \text{for all } (\pi_1,\pi_2) \in H_1\times H_2 \text{ and } s \in S.$$

Furthermore, for a game Γ with a state s and a pair of actions $(i,j) \in A_1(s)\times A_2(s)$, such that

$$p(s|s,i,a_2) = p(s|s,a_1,j) = 1 \quad \text{for all } a_1 \in A_1(s) \text{ and } a_2 \in A_2(s)$$

and such that

$$\sup_{a_1\in A_1(s)} r(s,a_1,j) = \inf_{a_2\in A_2(s)} r(s,i,a_2) = r(s,i,j),$$

assumption M guarantees that for the specific game with state s as starting state, the strategies "always playing action i", notation π_1^{is}, respectively "always playing action j", notation π_2^{js}, are optimal for player 1, respectively player 2. Then the value of the specific game with state s as starting state equals $w(\pi_1^{is},\pi_2^{js})(s)$.

Now, for a monotone evaluation rule w, let $G_S(w)$ be the family of games $\Gamma \in G_S$, which possesses a finite value with respect to the evaluation rule w. For an $s \in S$ and an $i \in A_1(s)$ let us denote by $\Pi_1(s,i)$ the family of history-dependent strategies $\pi_1 \in \Pi_1$, which do not use action i, i.e. if the system is at stage t in state s, then for each history of the game at state t, such a strategy π_1 selects a probability measure on $A_1(s)$ with mass zero in i.

<u>DEFINITION 4.3</u>. For a game $\Gamma \in G_S(w)$ an action $i \in A_1(\tilde{s})$ is non-essential for player 1, if for each $\tilde{\pi}_1 \in H_1$ and each $\varepsilon > 0$, there exists a strategy $\pi_1(\varepsilon,\tilde{\pi}_1) \in \Pi_1(\tilde{s},i)$, such that

$$w(\pi_1(\varepsilon,\tilde{\pi}_1),\pi_2)(s) \geq w(\tilde{\pi}_1,\pi_2)(s) - \varepsilon \quad \text{for all } \pi_2 \in H_2 \text{ and all } s \in S.$$

An action $j \in A_2(\tilde{s})$ is non-essential for player 2 if for each $\tilde{\pi}_2 \in H_2$ and each $\varepsilon > 0$, there exists a strategy $\pi_2(\varepsilon,\tilde{\pi}_2) \in \Pi_2(s,j)$, such that

$$w(\pi_1,\pi_2(\varepsilon,\tilde{\pi}_2))(s) \leq w(\pi_1,\tilde{\pi}_2)(s) + \varepsilon \quad \text{for all } \pi_1 \in H_1 \text{ and all } s \in S.$$

It can be seen that a game which results from a game in $G_S(w)$ after deleting a non-essential action for a player, is again a member of $G_S(w)$ and that the value does not change. So, in some sense, the set $A_1(s) - \{i\}$ is sufficient for player 1 if action i is non-essential.

In the next theorem we characterize the value function on the family $G_S(w)$. A proof can be obtained by slightly modifying the proof of Theorem 3.1 and will be omitted.

THEOREM 4.4. *A function* $f: G_S(w) \to \mathbb{R}^S$ *equals the value function if and only if* f *obeys the following four axioms:*

(1) axiom of objectivity: *if* $\Gamma \in G_S(w)$ *is such that for a state* s *both players have only one action, say* i *and* j *respectively, and if* $p(s|s,i,j) = 1$, *then* $f(\Gamma)(s) = w(\pi_1^{is},\pi_2^{js})(s)$.

(2) axiom of monotony: *if* $\Gamma' \leq \Gamma''$, *then* $f(\Gamma') \leq f(\Gamma'')$.

(3.i) axiom of sufficiency for player i; i = 1,2: *if* Γ' *is derived from* Γ *by deleting a certain state a non-essential action of player* i, *then* $f(\Gamma')(s) = f(\Gamma)(s)$ *for all* $s \in S$.

Similarly, as in Section 3, axiom 3.b can be replaced by a symmetry axiom and also in this case one can show that the axioms in Theorem 4.4 are independent.

REFERENCES

[1] DENARDO, E. & B. FOX, *Multichain Markov renewal programs*, SIAM J. Appl. Math. 16 (1968) 468-489.

[2] SHAPLEY, L.S., *Stochastic games*, Proc. Nat. Acad. Sci. USA 39 (1953) 1095-1100.

[3] TIJS, S.H., I: *A characterization of the value of linear programs*, II: *A characterization of the value of zero-sum two-person games*, Report 7632, Dept. of Math., Catholic University, Nijmegen, The Netherlands (1976).

[4] VILKAS, E.I., *Axiomatic definition of the value of a matrix game*, Theor. Prob. Appl. 8 (1963) 304-307.